Food and Agriculture Organization of the United Nations

Technical meeting on the Impact of Whole Genome Sequencing on food safety management

Poster abstracts

23 - 25 May 2016

Food and Agriculture Organization of the United Nations
Rome, Italy

Table of Contents

Executive summary

On the occasion of the Technical Meeting on the impact of Whole Genome Sequencing (WGS) on food safety management and the 9th meeting of the Global Microbial Identifier from 23 - 25 May 2016 in Rome, Italy, a poster session is held as an opportunity for experts to share and discuss their projects on the topic of WGS and food safety. The abstracts of all presented posters are compiled in this document. The topics of the posters range from research projects on WGS of specific microorganisms, developing national strategy to bring pathogen genomics into practice, different platforms for detection and analysis of foodborne outbreaks, to use of WGS in a factory environment. The poster session serves as a good platform for the participants of the Technical Meeting for further discussion and exchange of ideas on the topic of WGS for food safety.

Key words
Whole genome sequencing, WGS, microorganisms, pathogen, antimicrobial resistance, AMR, food safety, detection, analysis, foodborne outbreaks, poster, FAO

Year of publication: 2016

Background

On the occasion of the Technical Meeting on the impact of Whole Genome Sequencing on food safety management on 23-25 May 2016 in Rome, Italy, poster session is held as an opportunity for experts to share and discuss their projects on the topic of WGS and food safety. The posters are displayed in front of the David Lubin Library on the ground floor in the Building A. Posters are available during the meeting days and presenters are available for questions at 10.00 – 10.30 hrs every day.

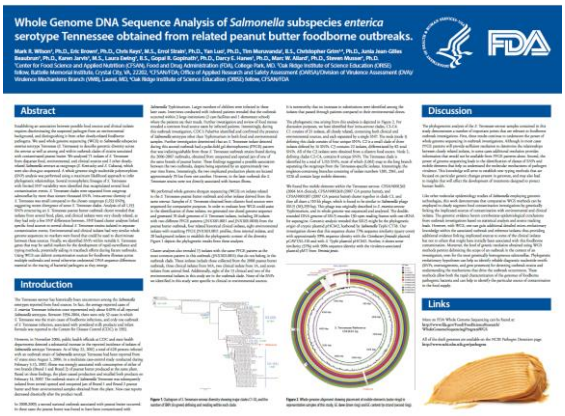
Poster Session

Venue: David Lubin Library (Ground Floor, Building A)
Dates: 23-25 May 2016
Time: All day
Presenters available for Questions: 10.00 – 10.30 hrs every day

Posters

1. Whole Genome Sequencing of *Salmonella* subspecies *enterica* serotype Tennessee obtained from peanut butter outbreaks

Marc Allard, US FDA CFSAN



Background

Establishing an association between possible food sources and clinical isolates requires discriminating the suspected pathogen from an environmental background, and distinguishing it from other closely-related foodborne pathogens.

Methods

We used whole genome sequencing (WGS) to *Salmonella* subspecies *enterica* serotype Tennessee (*S.* Tennessee) to describe genomic diversity across the serovar as well as among and within outbreak clades of strains associated with contaminated peanut butter. We analyzed 71 isolates of *S.* Tennessee from disparate food, environmental, and clinical sources and 2 other closely-related *Salmonella* serovars as outgroups (*S*. Kentucky and *S*. Cubana), which were also shot-gun sequenced. A whole genome single nucleotide polymorphism (SNP) analysis was performed using a maximum likelihood approach to infer phylogenetic relationships.

Results

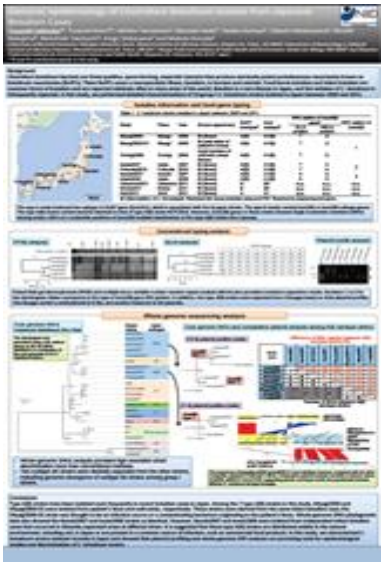
Several monophyletic lineages of *S*. Tennessee with limited SNP variability were identified that recapitulated several food contamination events. *S*. Tennessee clades were separated from outgroup salmonellae by more than sixteen thousand SNPs. Intra-serovar diversity of *S*. Tennessee was small compared to the chosen outgroups (1,153 SNPs), suggesting recent divergence of some *S*. Tennessee clades. Analysis of all 1,153 SNPs structuring an S. Tennessee peanut butter, outbreak cluster revealed that isolates from several food, plant, and clinical isolates were very closely related, as they had only a few SNP differences between.

Discussion

Single Nucleotide Polymorphism (SNP) -based cluster analyses linked specific food sources to several clinical S. Tennessee strains isolated in separate contamination events. Environmental and clinical isolates had very similar whole genome sequences; no markers were found that could be used to discriminate between these sources. Finally, we identified SNPs within variable *S*. Tennessee genes that may be useful markers for the development of rapid surveillance and typing methods, potentially aiding in traceback efforts during future outbreaks. Using WGS can delimit contamination sources for foodborne illnesses across multiple outbreaks and reveal otherwise undetected DNA sequence differences essential to the tracing of bacterial pathogens as they emerge. More on FDA Whole Genome Sequencing can be found at http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ and http://www.ncbi.nlm.nih.gov/pathogens/

2. Genomic epidemiology of *Clostridium botulinum* isolates from Japanese Botulism Cases

Tsuyoshi Sekizuka & Makoto Kuroda, Laboratory of Bacterial Genomics, Pathogen Genomic Center, National Institute of Infectious Disease



Background

Clostridium botulinum bacteria are Gram-positive, spore-forming, anaerobic bacteria that produce extremely potent proteinaceous neurotoxins known as botulinum neurotoxins (BoNTs). These BoNTs cause a neuroparalytic illness, botulism, in humans and animals. Food-borne botulism and infant botulism are common forms of botulism and are reported relatively often in many areas of the world. Botulism is a rare disease in Japan, and the isolation of *C. botulinum* is infrequently reported. In this study, we performed detailed characterizations of 10 group I *C. botulinum* strains isolated in Japan between 2006 and 2011.

Methods

Subtyping of BoNTs gene (*bont*) was examined by sequencing. Pulsed-field gel electrophoresis (PFGE) and Multiple-locus variable-number tandem repeat analysis (MLVA) was used as conventional methods for *C. botulinum* genotyping. Whole genome sequencing for *C. botulinum* strains was performed using an Illumina Genome Analyzer IIx and TruSeq SBS kit version 5 or an Illumina MiSeq and MiSeq reagent kit version 2 (Illumina) according to the manufacturer's instructions. Short-reads of all *C. botulinum* strains were compared against the reference chromosome sequence of *C. botulinum* ATCC 3502, followed by identification of single-nucleotide variation (SNV). All SNVs were concatenated to generate a pseudosequence for phylogenetic analysis; the maximum-likelihood-based DNA analysis program (RAxML version 7.25) was used for phylogenetic analysis with 1,000-times bootstrapping.

Results

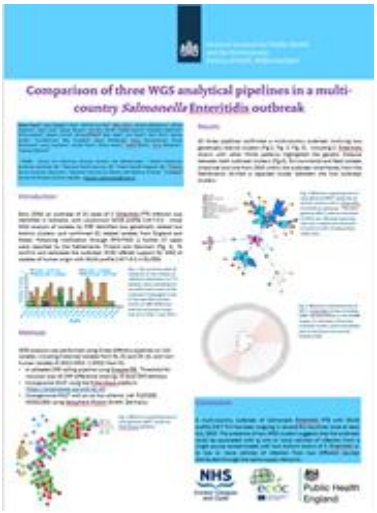
Genetic characterization was performed for 10 group I *C. botulinum* strains isolated from botulism cases in Japan between 2006 and 2011. Of these, 1 was type A, 2 were type B, and 7 were type A(B) {carrying a silent *bont/B* [*bont/(B)*] gene} serotype strains, based on botulinum neurotoxin (BoNT) production. The type A strain harbored the subtype A1 BoNT gene (*bont/A1*), which is associated with the *ha* gene cluster. The type B strains carried *bont/B5* or *bont/B6* subtype genes. The type A(B) strains carried *bont/A1* identical to that of type A(B) strain NCTC2916. However, *bont/(B)* genes in these strains showed SNVs among strains. SNVs at 2 nucleotide positions of *bont/(B)* enabled classification of the type A(B) strains into 3 groups. Pulsed-field gel electrophoresis (PFGE) and multiple-locus variable-number tandem-repeat analysis (MLVA) also provided consistent separation results. In addition, the type A(B) strains were separated into 2 lineages based on their plasmid profiles. One lineage carried a small plasmid (5.9 kb), and another harbored 21-kb plasmids. To obtain more detailed genetic information about the 10 strains, whole genome sequencing was performed. Whole-genome SNV analysis among 13 group I *C. botulinum* genomes provided high-resolution strain discrimination more than conventional methods and enabled to generate a refined phylogenetic tree that provides effective traceability of botulism cases. In the phylogenetic tree, two subtype B6 strains were distantly separated from the other strains, indicating genomic divergence of subtype B6 strains among group I strains.

Discussion and conclusions

Type A(B) strains have been isolated most frequently in recent botulism cases in Japan. Among the 7 type A(B) strains in this study, Miyagi2006 and Miyagi2006-01 were isolated from patient's feces and well water, respectively. These strains were derived from the same infant botulism case; the Miyagi2006-01 strain was thought to be an infection source or a contaminating bacterium originating in the patient's feces. Whole-genome SNV phylogenetic data also showed that the Ibaraki2007 and Iwate2008 strains were identical. However, Ibaraki2007 and Iwate2008 were isolated from independent infant botulism cases that occurred in distantly separated areas at different times. It is suggested that these type A(B) strains are distributed widely in the natural environment, including soil, in Japan or are present in a common source of infection, such as commercial food products. In this study, we characterized *C. botulinum* strains isolated recently in Japan and showed that small-plasmid profiling and whole-genome SNP analyses are promising tools for epidemiological studies and discrimination of *C. botulinum* strains.

3. Comparison of three WGS analytical pipelines in a multi-country Salmonella Enteritidis outbreak

Eelco Franz, National Institute for Public Health and the Environment (RIVM)



Background

On 18 January 2016, Health Protection Scotland (HPS) reported through the Epidemic Intelligence Information System for Food- and Waterborne Diseases and Zoonoses (EPIS-FWD) hosted at the European Centre for Disease Prevention and Control (ECDC) 21 cases of *Salmonella enterica* serovar Enteritidis PT 8 sharing an uncommon MLVA profile (2-9-7-3-2). The cases had arisen since late August 2015 and the number of cases represented an increase in comparison with previous years. Prior to this event this MLVA profile had last been seen in Scotland in August 2014. Based on this signal, other EU Member States reported human cases with the same MLVA profile. The Netherlands reported 15 cases sharing the same MLVA profile between May and December 2015, Finland reported one case from September 2015 and Denmark one case from August 2015. During preliminary analysis of the sequences from the Scottish isolates by Public Health England (PHE), an additional 52 isolates from England and Wales were identified, distributed in two distinct clusters. To confirm and delineate the outbreak, ECDC offered support for whole genome sequencing of human isolates with the MLVA profile 2-9-7-3-2 detected in any of the EU/EEA Member States. Isolates were accepted for sequencing independently on when they were collected.

Methods

For the purpose of outbreak definition the PHE single nucleotide polymorphism (SNP) approach was applied. Isolates were considered to be confirmed as part of the outbreak if they belonged in one of the two WGS clusters within ≤5 SNP difference (e.g. sharing the identical t5 level SNP address), and the symptoms onset was on or after 1 July 2015. Altogether, 104 isolates from 4 countries were included in the WGS analysis which was performed using three different pipelines; in the Netherlands an *ad hoc* core genome MultiLocus Sequence Typing (cgMLST) approach using SeqSphere software was used, the open access enterobase cgMLST scheme (http://enterobase.warwick.ac.uk/) was used in Scotland, and England and Wales used the SNP pipeline, already validated in PHE.

Results

Based on the results obtained using the PHE SNP pipeline, thirty-five isolates shared the t5 level SNP address 1.2.3.175.175.175.%, including 28 cases within the relevant time period (1July onwards) from the United Kingdom, two cases from the Netherlands and one case from Finland. Six additional historical isolates from the United Kingdom also belonged to this cluster. Fifty-two isolates shared the t5 level SNP address 1.2.3.18.359.360.% including 37 confirmed isolates within the relevant time period from the United Kingdom and seven confirmed isolates from the Netherlands. Fifteen historical isolates from cases before the outbreak period from the United Kingdom also belonged to this cluster. Altogether 66 cases from Finland, the Netherlands and the United Kingdom were classified as confirmed as part of the outbreak using WGS. Overall, all three analytical approaches confirmed a multi-country outbreak including two distinct clusters. Eight environmental and feed isolates from the Netherlands, including one from 2015, were not placed in either outbreak clusters, but formed a separate cluster in between the two outbreak clusters.

Discussion and conclusions

A multi-country outbreak of *Salmonella* Enteritidis PT8 with MLVA profile 2-9-7-3-2 has been ongoing in several EU countries since at least July 2015. Sixty-six cases from Finland, the Netherlands and the United Kingdom have been classified as outbreak cases on the basis of WGS analysis; eleven additional cases from Denmark, the Netherlands and the United Kingdom were classified as probable on the basis of MLVA analysis. The presence of two distinct WGS clusters with a similar distribution in time and place suggests that there are potentially two outbreaks or alternatively a single outbreak which could be associated with two or more vehicles of infection, although it is plausible that these could be contaminated by the same source. All three analytical pipelines used confirmed the outbreak and showed that human isolates were divided into two closely related genetic clusters. Some differences were observed between the results from the three analytical pipelines tested. These can be partly explained by the different approaches used but also by the application of different inclusion criteria for sequence quality. Participants of this case study have suggested to perform a more systematic comparison between the analytical pipelines in the near future.

4. Developing a national strategy to bring pathogen genomics into practice

Sobia Raza, PHG Foundation

Background

The PHG Foundation is an independent health policy think-tank with the mission to achieve responsible and evidence based application of biomedical science for health. The Foundation has been engaged on a major project spanning 18 months to assess the potential impact of microbial genomics on public health and health services and to develop policy to support implementation within England. The goal of this work was to identify and describe the actions that need to be taken to meet the aim of: '*Supporting the development and delivery of genomics informed infectious disease services that are evidence based, high quality, available population-wide, and on an equitable basis*'.

Methods

We undertook a programme of in-house research and analysis along with extensive stakeholder consultation over two workshops and also through one-to-one or small group engagement with academic researchers, clinical and public health service practitioners, and policy makers. We synthesised our findings into a comprehensive report, in addition to a series of policy briefings focussed on different aspects of pathogen genomics implementation, such as disease surveillance and ensuring a 'one health' approach to infectious disease management.

Results

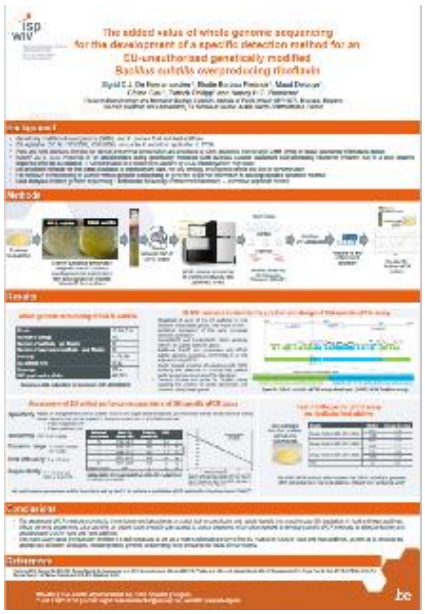
Our report reviews the current state of science and clinical practice in pathogen genomics, and details 35 recommendations to achieve the above goal. Our results clearly show that the effectiveness of any efforts to implement pathogen genomics will depend on the ability to realize a nationally coordinated system of service development and delivery. Two essential features of this system are: Strategic coordination and leadership: across the organizations involved in delivering and using pathogen genomics. The management of infectious disease and its impact on human health in England requires the input of a wide range of organizations including those with responsibility for public health, clinical healthcare, food safety, and animal health, as well as a wide range of professional groups. Since each of these organizations and professional groups has a stake in realizing the effective development and implementation of pathogen genomics services, their efforts can only succeed where there are clear mechanisms to achieve strategic coordination of policy at an organizational level and where there are mechanisms to ensure that professional groups are supported to work together to share and develop the knowledge, expertise and best practice.
Data integration: of all genomic and accompanying metadata. The benefits of pathogen genomics can only fully be realized through the timely collation, integration and sharing of genomic and clinical/epidemiological metadata by those involved in the delivery of pathogen genomics services. Data sharing and integration across sectors (human, animal, food, environment) will not only serve to delivering improvements in outbreak detection and resolution, but is also essential for driving innovation and the expansion of services for future use.

Conclusions

Whilst our report focusses on the delivery of pathogen genomic services within the context of the English health system, the findings and recommendations are highly transposable to other health economies. Central to our recommendations is the call for all stakeholders to engage with addressing the challenge of adopting a 'whole system' approach that, just as pathogens cross barriers between species, will cut across national and international barriers and specialism and organizational silos. These features and others are discussed in depth in our report: *Pathogen Genomics into Practice*

5. The added value of whole genome sequencing for the development of a specific detection method for an EU-unauthorized genetically modified *Bacillus subtilis* overproducing riboflavin

Sigrid De Keersmaecker, Scientific Institute of Public Health, Belgium (WIV-ISP), Platform Biotechnology and Molecular Biology



Background

Genetically modified micro-organisms (GMM) are known to be used to produce food and feed additives. In the framework of the EU legislation (EC No 1333/2008, 1331/2008), companies aiming at introducing at the EU market a specific additive produced by GMMs, should submit an application to the European Food Safety Authority (EFSA). In September 2014, the presence of an unauthorized living genetically modified (GM) *Bacillus subtilis* bacterium overproducing vitamin B2 in a feed additive imported into the EU market was notified by the Rapid Alert System for Food and Feed (RASFF 2014–1249). This event has highlighted the fact that contamination of a food or feed additive by a GM-microorganism may occur. Moreover, it confronted for the first time the enforcement laboratories with this type of contamination, while at that time no detection method for this GMM was available, nor was there any strategy developed to detect this kind of contamination. For the specific case of the vitamin B2 overproducing *B. subtilis*, as no sequence information of this GMM nor any specific detection or identification method was available, whole genome sequencing was used to generate sequence information to be able to develop such specific detection method. However, the data analysis of whole genome sequencing often requires bioinformatics expertise which is not always present in the average enforcement laboratory. This might impede the use of this powerful technology to rapidly obtain critical sequence information required to develop a specific detection method for GMM.

Methods

The French National Reference Laboratory (NRL) for GMO isolated a yellow substance (presumably overproduction of riboflavin)-secreting bacterial strain from three samples of imported Vitamin B2 feed additives. Subsequently, genomic DNA of one of the isolates was extracted and used for whole-genome sequencing on an Illumina HiSeq2500 run using a paired-end library. Sequencing reads were assembled de novo using the "De novo assembly" option of the CLC Genomics Workbench version 7.5.1. (CLC Bio). The resulting draft genome was further linked into scaffolds with SSPACE based on paired-end read linkage. Sequencing data (JYFL00000000) were analysed using a simple BLAST approach. Based on the results, a TaqMan® qPCR method was developed and tested on isolated bacterial strains and on the feed additive directly.

Results

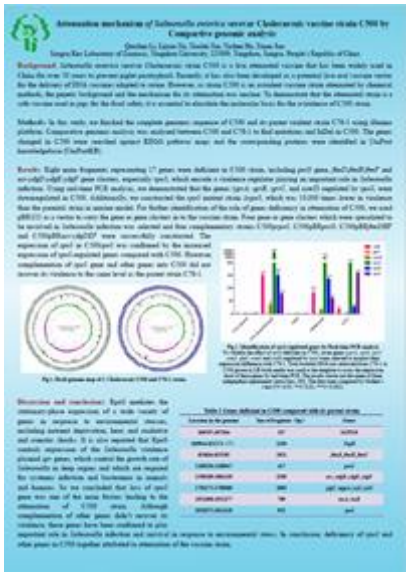
Whole genome sequencing yielded 10,914,314 paired-end reads (350-fold coverage), which were assembled de novo and put together into 36 *B. subtilis* specific scaffolds (JYFL00000000). The total sequence length was 4,175,764 bp, and had a G+C content of 44.32%. These scaffolds were used in a simple strategy based on the common BLAST tools. This approach can be used by any enforcement laboratory lacking deep bioinformatics expertise. The results were used to design and assess a new TaqMan® qPCR method (called "VitB2-UGM"), targeting the junction between the *B. subtilis* riboflavin operon and the vector used to construct this GM vitamin B2 overproducing bacterium. The method was developed according to EU critical performance parameters for specificity, sensitivity, PCR efficiency and repeatability. The VitB2-UGM method was also able to detect the *B. subtilis* bacterium in genomic DNA extracted from the feed additive, without prior culturing step.

Discussion and conclusions

The developed qPCR method has provided to the enforcement laboratories a crucial tool to specifically and rapidly identify this unauthorized GM bacterium in food and feed additives. Moreover, this study illustrates how the use of whole genome sequencing data can offer an added value to easily gain access to crucial sequence information needed to develop specific qPCR methods to detect unknown and unauthorized GMO in food and feed additives. In addition, this case study raises the question whether it is not necessary to set up a more systematic survey of the EU market for GMM in food and feed additives, as well as to develop the appropriate detection strategies (including whole genome sequencing) to be prepared for future similar events.

6. Attenuation mechanism of *Salmonella enterica* serovar Choleraesuis vaccine strain C500 by Comparative genomic analysis

Qiuchun Li & Xiaolei Xie, Yangzhou University



Background

Salmonella enterica serovar Choleraesuis strain C500 is a live attenuated vaccine that has been widely used in China for over 50 years to prevent piglet paratyphoid. Recently, it has also been developed as a potential live oral vaccine vector for the delivery of DNA vaccines adapted to swine. However, as strain C500 is an avirulent vaccine strain attenuated by chemical methods, the genetic background and the mechanism for its attenuation was unclear. To demonstrate that the attenuated strain is a safe vaccine used in pigs for the food safety, it is essential to elucidate the molecular basis for the avirulence of C500 strain.

Methods

In this study, we finished the complete genomic sequence of C500 and its parent virulent strain C78-1 using illunina platform. Comparative genomic analysis was analyzed between C500 and C78-1 to find mutations and InDel in C500. SNPs analysis between two strains were performed using the NUCmer and shown-snps programs in the NUMmer 3 package.The SNPs located within coding regions causing synonymous or nonsynonymous were identified by SNPA program. The genes changed in C500 were searched against KEGG pathway maps and the corresponding proteins were identified in UniProt knowledgebase (UniProtKB).

Results

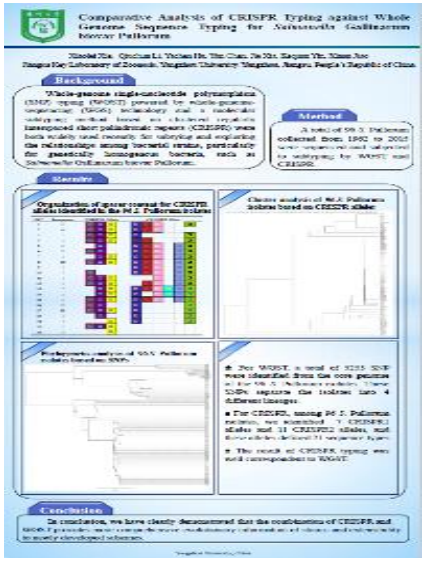
Eight main fragments representing 17 genes were deficient in C500 strain, including *ptsG* gene, *fimD-fimH-fimF* and *asr-ydgD-ydgE-ydgF* gene clusters, expecially *rpoS*, which encode a virulence regulator playing an important role in *Salmonella* infection. Using real-time PCR analysis, we demonstrated that the genes (*spvA, spvB, spvC,* and *osmY*) regulated by *rpoS*, were downregulated in C500. Additionally, we constructed the *rpoS* mutant strain Δ*rpoS*, which was 10,000 times lower in virulence than the parental strain in murine model. For further identification of the role of genes deficiency in attenuation of C500, we used pBR322 as a vector to carry the gene or gene clusters in to the vaccine strain. Four gene or gene clusters which were speculated to be involved in *Salmonella* infection was selected and four complementary strains C500p*rpoS*, C500pBR*ptsG*, C500pBR*fimDHF* and C500pBR*asr-ydgDEF* were successfully constructed. The expression of *rpoS* in C500*rpoS* was confirmed by the increased expression of *rpoS*-regulated genes compared with C500. However, complemention of *rpoS* gene and other genes into C500 did not recover its virulence to the same level as the parent strain C78-1.

Discussion and conclusions

RpoS mediates the stationary-phase expression of a wide variety of genes in response to environmental stresses, including nutrient deprivation, heat, and oxidative and osmotic shocks. It is also reported that RpoS controls expression of the *Salmonella* virulence plasmid *spv* genes, which control the growth rate of *Salmonella* in deep organs and which are required for systemic infection and bacteremia in animals and humans. So we concluded that loss of *rpoS* gene was one of the main factors leading to the attenuation of C500 strain. Although complemention of other genes didn't recover its virulence, but these genes have been confirmed to play important role in *Salmonella* infection and survival in response to environmental stress. In conclusion, deficiency of *rpoS* and other genes in C500 together attributed to attenuation of the vaccine strain.

7. Comparative Analysis of CRISPR Typing against Whole Genome Sequence Typing for *Salmonella* Gallinarum biovar Pullorum

Xiaolei Xie & Qiuchun Li, Yangzhou University



Background

Whole-genome single-nucleotide polymorphism (SNP) typing (WGST) powered by whole-genome-sequencing (WGS) technology and a molecular subtyping method based on clustered regularly interspaced short palindromic repeats (CRISPR) were both widely used recently for subtying and exploring the relationships among bacterial strains, particularly for genetically homogenous bacteria, such as *Salmonella* Gallinarum biovar Pullorum.

Whole Genome Sequence Typing (WGST) is a method that maximizes the data available for inference of genetic diversity and has been successfully used to distinguish among highly related isolates. With WGST, genetic relationships among isolates are determined by the number of single-nucleotide polymorphism (SNP), usually, the fewer the number of SNPs observed between the strains, the more closely the strains are related.

CRISPR loci are arrays of short repeats separated by short spacers. Along with the CRISPR associated (*cas*) genes, they provide an adaptive immunity for archaea and bacteria to defend against foreign invading genetic elements. Through the acquisition of new spacers and deletion of old spacers, differences useful for typing could occurred even between genetically homogeneous strains. Thus, Spacers may serve as a biogeographic marker used for differentiating individuals from different environments or sources.

Methods

In order to compare the subtyping result between CRISPR typing and WGST, a total of 96 *S*. Pullorum collected from 1962 to 2015 were sequenced and subjected to subtyping by these two methods. We sequenced 96 strains of *salmonella* Pullorum, and mapped the paired-end reads to the reference genome RKS5078. A cladogram was generated by using 5253 SNP sites which divided Pullorum into 4 clades. For each sequenced locus, CRISPR1 and CRISPR2 spacers were found by CRISPR-finder (http:// crispr.u-psud.fr/Server/) and were visualized as described by Deveau et al. Cluster analyses were performed based on allelic profile data by the unweighted pair group method with arithmetic mean (UPGMA).

Results

For WGST, a total of 5253 SNP were identified from the core genome of the 96 *S*. Pullorum isolates. These SNPs separate the isolates into 4 different lineages, and the largest lineage consists of 67 isolates, which accounts for 70% of all isolates. For CRISPR, among 96 *S*.Pullorum isolates, we identified 8 CRISPR1 alleles and 14 CRISPR2 alleles, and these alleles defined 23 sequence types. Most isolates (46, 48%) belong to the same CRISPR type, named PST 1(Pullroum sequence type). It is worth mentioning that almost all the isolates of PST1 except one belong to the largest lineage of WGST.

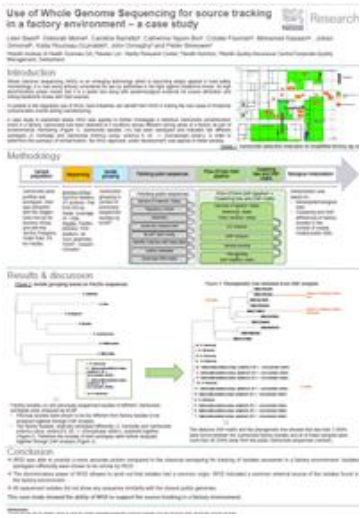
Discussion and conclusions

In this study, we found that both WGST and CRISPR typing were efficient and powerful to distinguish strains isolated from different sources or years, and the result of CRISPR typing was well correspondent to WGST. CRISPR sequence data shows subtle but distinct differences among different strains, and the dynamic adaptive nature of CRISPR loci provide critical insights into the relationships between strains and their environments. But, it is not comprehensive enough to show the differentiation on the whole genome level. While, WGST based on WGS can provide us with more comprehensive information on the standard of whole genome, but it is difficult to explain the subtle difference between isolates.

It is not a stretch to infer that typing methods based on WGS will eventually become the new gold standard for microbial pathogen subtyping. In conclusion, we have clearly demonstrated that the combination of CRISPR and WGST provides more comprehensive evolutionary information of strains and extensibility to newly developed schemes.

8. Use of Whole Genome Sequencing for source tracing in a factory environment – a case study

Leen Baert & Caroline Barretto, Nestle Research Center



Background

Whole Genome Sequencing (WGS) is an emerging technology which is becoming widely applied in food safety microbiology. It is now being actively considered for use by authorities in the fight against foodborne illness. Its high discriminatory power means that it is a useful tool along with epidemiological evidence for source attribution and linking foodborne illness with food sources. In parallel to the regulatory use of WGS, food industries can benefit of WGS in finding the root cause of incidental contamination events during manufacturing. A case study is presented where WGS was applied to further investigate a historical *Salmonella* contamination event in a factory. *Salmonella* had been detected at different locations across different zoning areas of a factory as part of environmental monitoring. *Salmonella* isolates had been serotyped and indicated two different serotypes (*S.* Kentucky and *Salmonella enterica subsp. enterica* 8, 20 : i:- (monophasic strain)). In order to determine the source(s) of contamination, the WGS approach, under development, was applied to these isolates.

Methods

Salmonella were cultivated, isolated and purified. DNA was extracted and processed for sequencing on Illumina MiSeq or HiSeq platforms. Prior to sequencing, libraries were prepared following Illumina Nextera XT protocol. The libraries were quality checked and quantified using capillary electrophoresis. The samples were pooled together with a close to an equimolar concentration and sequenced with Pair End 250 bp reads at a coverage of approximately 100x.

In parallel, few selected isolates were also sequenced using Pacific Biosciences long read sequencing technology (PacBio) to create reference genomes. 20 kb sequencing libraries were built following PacBio protocol and sequenced on a RSII platform. Afterwards the bacterial genomes were *de novo* assembled using the Hierarchical Genome Assembly Process (HGAP) software, followed by genome closure using Circulator. The quality of the assemblies was validated using a quality verification pipeline (Mega BLAST and dot plot comparisons).

Bioinformatics analysis of the factory contamination consisted in clustering the isolates (kSNP) and then performing the in depth SNP analysis with the adapted CFSAN-FDA pipeline including also the closest public genomes available to put the results into a biological context. A selected high quality genome generated with PacBio was used as a reference for SNP calling based on whole genome read mapping of Illumina short reads. The bioinformatics pipeline was set up as the following: (I) mapping of reads from the samples on the reference genome, (II) SNP identification and filtering, (III) alignment of all SNP identified, (IV) Phylogenic tree and SNP distance matrix, (V) interpretation of results.

Results

The genomes sequenced with PacBio were successfully assembled into a single large contig representing the chromosome.The factory isolates, originally serotyped differently (*S.* Kentucky and *Salmonella enterica subsp. enterica* 8, 20 : i:- (monophasic strain)), clustered together through kSNP. Therefore the isolates of both serotypes were further analyzed together through SNP analysis. The distance SNP matrix and the phylogenetic tree showed that less than 5 SNPs were found between the *Salmonella* factory isolates and all samples were more than 50 SNPs away from the public *Salmonella* sequences. The interpretation of the results was that all factory isolates were coming from the same origin and didn't show sequence similarity with the public genomes.

Discussion and conclusions

Whole Genome Sequencing (WGS) was able to provide a more accurate picture compared to the classical serotyping for tracking of isolates recovered in a factory environment. Isolates serotyped differently were shown to be similar by WGS. The discriminatory power of WGS allowed to point out that isolates had a common origin. WGS indicated a common external source of the isolates found in the factory environment.

This case study showed the ability of WGS to support the source tracking in a factory environment.

9. Collaborative Management Platform for Detection and Analyses of (Re)-emerging and Foodborne Outbreaks in Europe

Rene Hendriksen , Technical University of Denmark & Marion Koopmans, , Erasmus MC

Background

COMPARE aims to harness the rapid advances in molecular technology to improve identification and mitigation of emerging infectious diseases and foodborne outbreaks. COMPARE has brought together 29 partners to develop a global platform for exchange of data and analyses. This Horizon 2020 project will run between December 2014 and November 2019

Objectives

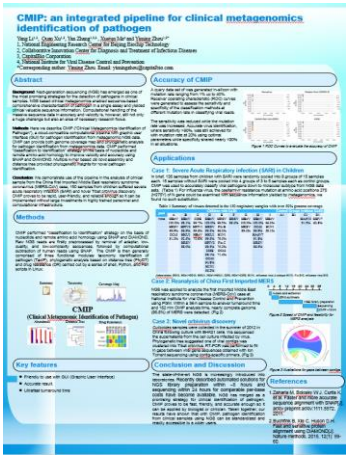
To improve rapid identification, containment and mitigation of emerging infectious diseases and foodborne outbreaks
To develop a cross-sector and cross-pathogen analytical framework and globally linked data- and information-sharing platform
To integrate state-of-the-art strategies, tools, technologies, and methods for collecting, processing and analysing sequence-based pathogen data in combination with associated data
To generate actionable information for relevant authorities and other users in the human health, animal health and food safety domains

Features Available

Reference genome database posted on European Nucleotide Archive (ENA).
COMPARE data hubs launched, allowing partners to share data rapidly in a structured way with the possibility to 'quarantine' data.
COMPARE virtual machine (VM), a cloud compute environment, bringing computing power and the data together.
Trial phase for COMPARE 'Notebook' environment; COMPARE bioinformaticians can do ad hoc analysis tasks.

10. CMIP: an integrated pipeline of clinical metagenomics for pathogen identification

Yang Li & Yiming Zhou, National collaborative innovation center for diagnosis and treatment of infectious diseases



Background

Next-generation sequencing (NGS) has emerged as one of the most promising strategies for the detection of pathogens in clinical samples. NGS based clinical metagenomics enabled sequence-based comprehensive characterization of pathogen in a single assay and yielded clinical valuable sequence information. However, Computational handling of the massive sequence data in accuracy and velocity is, however, still not only a huge challenge but also an area of necessary research focus. It is noted that most viral pathogens are RNA viruses which suggested the divergent genomes. It could be impossible to find a related reference genome during alignment procedure in some cases. The construction of a phylogenetic tree allows us to visualize the underlying genealogy between the divergent virus and existing reference sequences.

Methods

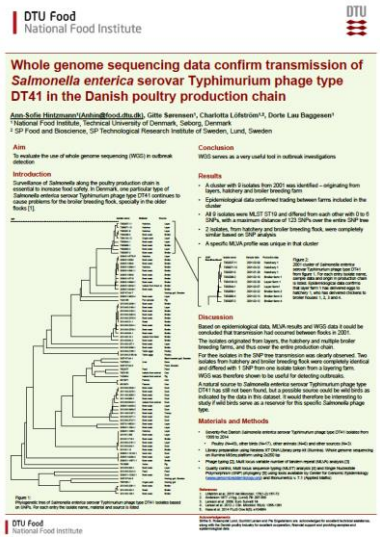
Here we describe CMIP ("Clinical Metagenomics Identification of Pathogen"), a cloud-compatible computational pipeline with graphic user interface (GUI) for pathogen identification from metagenomic NGS data. CMIP can provide both genome coverage map and phylogenetic analysis for pathogen identification from metagenomics data. CMIP performed "classification to identification" strategy on the basis of nucleotide and remote amino acid homology to improve velocity and accuracy using SNAP and DIAMOND. Multiple k-mer based *de novo* assembly and distance tree provided phylogenetic insights for novel pathogen identification.

Conclusion

We demonstrate use of the pipeline in the analysis of clinical sample from the China first imported Middle East respiratory syndrome coronavirus (MERS-CoV) case, 150 samples from children suffered severe acute respiratory infection and novel Tibet orbivirus discovery. CMIP proves to be rapid, user-friendly, and reliable enough so it can be implemented without large investments in highly trained personnel and computational infrastructure.

11. Whole genome sequencing data confirm transmission of Salmonella enterica serovar Typhimurium phage type DT41 in Danish poultry production

Charlotta Löfström, SP Technical Research Institute of Sweden


DTU Food
National Food Institute

Whole genome sequencing data confirm transmission of *Salmonella enterica* serovar Typhimurium phage type DT41 in the Danish poultry production chain

Aim

Introduction

Conclusion

Results

Discussion

Materials and Methods

DTU Food
National Food Institute

Background

Salmonella causes numerous foodborne outbreaks every year in both developing and developed countries. It spreads through the entire food production chain and thus can be difficult to manage. Consumption of contaminated poultry is one of the main sources of human salmonellosis and it is therefore important to have efficient methods for detection and characterization of Salmonella isolates, e.g. in surveillance systems and for outbreak investigations, to increase food safety. One specific type Salmonella enterica serovar Typhimurium phage type DT41 (S. Typhimurium DT41) has previously been isolated on repeated occasions from broiler breeder flocks in Denmark, particularly in older poultry. The aim of this study was to investigate the usefulness of whole genome sequencing (WGS) to follow transmission of Salmonella in the Danish poultry production chain, with focus on primary production.

Methods

Danish Salmonella ser. Typhimurium DT41 isolates from 1999 to 2014 covering poultry (n = 50), other birds (n = 17), other animals (n = 6) and other sources (n = 3) were analysed with WGS (library preparation using Nextera XT DNA Library prep kit (Illumina) and WGS on Illumina MiSeq platform using 2x250 bp). WGS results were compared to results obtained using Multi Locus Variable number tandem repeats Analysis (MLVA) results. Isolates were either epidemiologically linked or considered as "background" to increase the resolution of the analysis. The Multi Locus Sequence Type (MLST) and Single Nucleotide Polymorphism (SNP) phylogeny was constructed using tools available at the Center for Genomic Epidemiology (www.genomicepidemiology.org) and BioNumerics v. 7.1 (Applied Maths).

Results

A SNP tree was constructed based on results from the WGS analysis, where a cluster with nine isolates from 2001 was identified. This cluster contained isolates originating from layers, hatchery and broiler breeding farm. Epidemiological data confirmed trading between farms included in the cluster and thus a possible transmission from layers to hatchery and broiler breeding farm. All isolates in the cluster had the same MLST type (ST19) and had a minimum and maximum distance of 0 and 6 SNPs, respectively, and a max distance of 123 SNPs over the entire tree. The isolates shared a common MLVA profile that was otherwise not seen in the entire dataset, confirming a link between the isolates. Within the cluster there were two identical isolates, one is from a hatchery and the other is from a broiler breeding flock, suggesting that a direct transmission event has occurred.

Discussion and conclusion

Data from this study shows that WGS can be a useful method to solve outbreaks and study transmission routes in the poultry production chain. A natural source of the Salmonella enterica serovar Typhimurium phage type DT41 outbreak in Danish poultry production has still not been found, but a possible source could be wild birds as indicated by results from this study. More research is therefore planned in an attempt to identify a source to this specific Salmonella phage type in Danish poultry production. Furthermore, the applicability of WGS for outbreak investigations and study of contamination through food production will be included in future studies