final report

Project code: V.MFS.0415

Prepared by: P. Scott Chandry
Commonwealth Scientific and Industrial Research Organisation,
Agriculture and Food

Date published: 5 March 2018

PUBLISHED BY
Meat and Livestock Australia Limited
Locked Bag 1961
NORTH SYDNEY NSW 2059

Developing an industry strategy for use of new genetic identification systems and surveillance technologies

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government and contributions from the Australian Meat Processor Corporation to support the research and development detailed in this publication.

Abstract

This report constitutes one of two documents prepared for MLA Project No. V.MFS.0415

Part 1, "Developing an industry strategy for use of new genetic identification systems and surveillance technologies". It provides background data on the application of next generation DNA sequencing (NGS) data to methods such as whole genome sequencing (WGS) relevant to the red meat industry. In addition to a description of the technology itself, alternative applications of NGS are outlined. Potential issues and benefits to the use of WGS for the analysis of microbes are provided. Finally, a series of recommendations for potential industry actions are also provided. Throughout this document plain English summaries are provided at the beginning of each section.

Part 2, titled Implementation of Next Generation Sequencing in Food Microbiology Testing in Australia has been prepared by the Queensland Department of Health. This report describes collated data from a phone survey conducted with broad range of major Australian laboratories and industry stakeholders involved in testing of microbial isolates from food products concerning the implementation and use of whole genome sequencing (WGS) for bacterial testing and typing.

1 Executive Summary

Technological and computational advances in next generation DNA sequencing (NGS) has led to whole genome sequencing (WGS) becoming a required application in most microbial research. These advances have not been limited to research, NGS is rapidly becoming the "gold standard" technology for public health and food regulatory agencies around the world. Therefore, the red meat industry will be both directly and indirectly affected by the changes brought about by DNA sequencing. This will range from the impact on the genetics of animals, to soil management, pasture and crop improvements and testing / control of microbial contamination. This document provides some background information to assist in the red meat industry's response to the development of new genetic identification and surveillance technologies for food safety aspects of public health prompted by the advances in DNA sequencing technology.

Whole Genome Sequencing is an improved microbial typing method with substantially greater power to discriminate between closely related bacteria. It represents a dramatic improvement over previous typing methods such as pulsed field gel electrophoresis (PFGE). The high level of precision in WGS data yields the most effective means of tracking outbreaks and identifying isolates. This technology is already in use around the world. The most mature WGS based system is the GenomeTrakr network in the US that has already been instrumental in regulatory interventions with food processors. GenomeTrakr captures data about microbes from the US and internationally derived samples. This includes Australian microbial isolates detected in the US and several Australian institutions that contribute WGS data to the publically available GenomeTrakr network. So, some Australian data is already publically accessible. An example of a system similar to that of the US but from a similar sized country to Australia is Canada that has a well-developed system using WGS data for food safety microbiology. The modular WGS based system implemented at the Canadian Food Inspection Agency reduces the time and cost of the analysis of microbial isolates while increasing the amount of data that is generated.

Some issues raised by uptake of WGS and NGS technologies in Australia includes the following:

- The need for new regulations and standards (for technology, methods, analysis, and data storage).
- The requirement for development of new testing regimes and naming standards to assure consistency of results between testing facilities.
- Allocation of the resources required to store large amounts of WGS and computer analysis data (in the cloud or locally).
- Fulfil the need for training of testing lab staff, quality control staff, and regulators to cope with new data types.
- Decisions concerning who will have access to WGS data (public vs. private).
- In Australia there is a minimal amount of WGS data for isolated microbes from food, processing facilities, and the environment. This in turn limits our knowledge of how diverse these microbes are which means it will be difficult to know if a microbe found by one producer is unique to a single facility (region, state) or if it is found everywhere.

- WGS data may permit Australia to capitalise on the limited risk posed by local pathogenic microbes (when compared to the same organisms from overseas).
- Australian regulators and health labs have begun adoption of NGS/WGS later than many parts of the world and this may limit the role Australia can play in how this technology is implemented in international standards and regulations.

The adoption of WGS and NGS methods to microbial surveillance and tracking could result in a number of potential benefits to the red meat industry such as:

- WGS is superior to all previous methods for the tracking and surveillance of microbial isolates leading to fewer mistakes in identification and attribution of microbes.
- The high sensitivity of WGS methods can permit the discrimination of local variations in microbes so it may be possible to tell if a bacterium is from producer A or producer B. *This is a simplified description since epidemiological investigations do not rely upon a single data source.*
- Implementation of WGS methods for the analysis of isolates should make analysis both quicker and cheaper.
- The information provided by WGS makes the detection of emerging microbiological threats easier since you do not have to know what you are looking for to detect something new.
- WGS provides a wealth of data about an isolate, making some additional testing redundant (e.g., antimicrobial resistance testing). Automated analysis packages now exist that can provide this information in a human readable form to end users.

2 Contents

Document Overview i
Part 1: Developing an industry strategy for use of new genetic identification systems and surveillance technologies 6
1 Background and Introduction 7
1.1 Why sequence DNA 8
1.2 What is NGS 9
1.3 Overview of NGS technologies 10
2 What can NGS data be used for 13
3 NGS/WGS for surveillance, outbreaks, and source tracking 16
3.1 The Canadian example for implementation of NGS/WGS 17
4 Issues caused by NGS and WGS 18
4.1 Local issues for the Australian Beef Industry 20
5 Benefits of WGS 22
6 Recommendations 24
7 Useful resources for additional information 26
8 References. 27
Part 2: Implementation of Next Generation Sequencing in Food Microbiology Testing in Australia 28
Tables 29
Glossary 30
Summary 31
1 Background 32
2 Objective 33
3 Survey Method 34
4 Survey Results 35
4.1 Current State 35
4.2 Data Access and Metadata association 35
4.3 Accreditation 36
4.4 Methodology 36
4.5 Turn Around Time 37
4.6 Future Directions 37
4.7 Open Comments 37
4.8 Themes identified during Interviews 38
Appendices 39
References 40
Appendix 1: Questions 41
Appendix 2: Organisations Contacted 42

Part 1: Developing an industry strategy for use of new genetic identification systems and surveillance technologies

3 Background and Introduction

Next Generations DNA Sequencing (NGS) which underpins Whole Genome Sequencing (WGS) is an improved microbial typing method with substantially greater power to discriminate between closely related bacteria. It represents a dramatic improvement over previous typing methods such as PFGE. The high level of precision of WGS data yields the most effective means of tracking outbreaks and identifying isolates.

Technological and computational advances in the sequencing of DNA has transformed most of the biological sciences. From the varied habitats studied in ecology to the minutia of genetics, all have been profoundly changed and microbiology is no exception to this trend. The impact of this transformation is not limited to scientific research, it is now impacting the entire health profession, agriculture, and many industries. Therefore, the meat industry will be both directly and indirectly affected by the changes brought about by DNA sequencing. This will range from the impact on the genetics of animals, to soil management, pasture and crop improvements and testing / control of microbial contamination. This document will attempt to provide some background information to assist in the red meat industry's response to deal with the development of new genetic identification and surveillance technologies prompted by the advances in DNA sequencing technology.

Since the first commercial next generation sequencing (NGS[1]) equipment became available (~2007), whole genome[2] sequencing (WGS[3]) has become a standard application in most microbial research. The smaller size of microbial genomes compared to those of more complex (so-called "higher") organisms has made application of NGS technology almost compulsory for microbiological research. These advances have not been limited to the realm of research, NGS is rapidly becoming the "gold standard" technology for public health and food regulatory agencies around the world (Nadon *et al.* 2017). The recent proliferation in the use of WGS for typing bacterial pathogens involved in food borne disease outbreaks in the USA, Canada, Europe and the UK indicates that it will become the standard technology for disease investigation globally. This technology will replace commonly used typing methods such as Pulsed Field Gel Electrophoresis (PFGE), Multi Locus Sequence Typing (MLST) and Multi Locus Variable number tandem repeat Analysis (MLVA).

Effective management of microbial pathogen contamination requires that microorganisms be identified beyond simple designations like species. For example, it is insufficient to establish that a microbe is an *Escherichia coli* (also known as *E. coli*), the type of *E. coli* is also important. An *E. coli* O157 is a pathogen classified as an adulterant when present in meat (in the USA) while *E. coli* Nissle 1917 is considered a probiotic and is a beneficial microorganism. Typing technologies provide a means to group microbes with more precision than simply the species, they provide an understanding of the relationship between different isolates of the same species. The method or methods used to type bacteria will determine the precision and certainty to which relationships can be inferred. One form of typing, serotyping, is based on the response caused by the binding of antibodies to structures on the surface of a microbial isolate. An example is important pathogen *E. coli* O157:H7 where the letters "O" and "H" describe type surface structures bound by test antibodies. In this case O antigen number 157 is a surface lipopolysaccharide that

[1] NGS – Next Generation Sequencing, a technology that generates millions of DNA sequencing reads per run compared to only hundreds of reads produced by the previous generation technology.

[2] Genome – the entirety of the heritable genetic material in a living organism. In the case of bacteria this would include the large chromosome (or chromosomes) as well as any smaller genetic material (e. g. plasmid) that is replicated and transferred from "mother" to "daughter" cells.

[3] WGS – Whole Genome Sequencing – reading every base (essentially every letter of the genetic code) of both the main chromosome(s) as well as any other genetic material that is replicated – see Genome description above

plays numerous roles in cell stickiness and protection from the immune system while the H antigen number 7 is the flagellum[4]. Although serotyping is a valuable means of categorizing microbes it lacks sufficient resolution to discriminate between different isolates of the same serotype. The tracking of pathogen outbreaks or contamination incidents requires a typing method that is reproducible in any lab and is capable of discriminating closely related microbial isolates of the same serotype. A wide variety of typing technologies were tested but the development of PFGE in 1984 advanced a novel method that became the gold standard for typing pathogen isolates. This development permitted unprecedented and highly reproducible discrimination of closely related isolates of a single species and serotype.

The specifics of PFGE are not crucial for this discussion but in essence it involves determining a "fingerprint" or pattern based on the size of cut pieces of genomic DNA. The size of the cut DNA fragments measures the presence of enzyme cut sites and the distance between each site. When two microbial isolates are the same or very similar then the frequency and position of these cut sites will be the same. The more closely related that organisms are the fewer variations will be observed in the number of sites and the distance that separates them. A prescribed methodology permits laboratories all over the world to compare results. The "fingerprint" patterns can be interrogated using the PulseNet database. Although PFGE is a reproducible and accurate means of typing closely related strains it has several limitations. Despite the existence of prescribed methods, PFGE results can vary between labs, fingerprint patterns cannot distinguish DNA fragments that are similar size but otherwise very different in composition, and most importantly the degree of discrimination between isolates is limited by the nature of the method itself. A typical PFGE preparation might cut a recognition site in the genome of an isolate approximately 25 times. This means approximately 150 bases were examined and a rough estimate of distance between these sequences was determined. This yields an accurate but limited analysis using very roughly 0.003 % of the information encoded in the genome (assuming 25 cuts, 6 bases per cut, and a genome of five million bases). NGS based methods are capable of analysing the millions of bases that compose the genome providing a far greater degree of discrimination between isolates than is possible with PFGE. If the genome of a microbial isolate were likened to a book PFGE essentially examines the length of each chapter to determine how similar the "books" are. NGS based methods such as WGS compare letter by letter and word by word eliminating most of the uncertainty inherent in PFGE.

3.1 Why sequence DNA

DNA is the instruction set for life so it underpins the definitive methods to identify and characterise all living things. Whole genome sequencing (reading all the DNA in an organism) will be the gold standard method for characterising and comparing microbes for the foreseeable future.

Continuing the analogy above comparing a microbe to a book, means the DNA of the genome are the words that describe the instructions for life itself, essentially a "book of life" for a particular organism. The words in this book contain only four possible letters A, C, G, and T but when arranged in various combinations it is the blue-print for constructing every physical structure that forms a microbial cell as well as the instruction set for how the microbe behaves in the environment. It determines where it will grow, what it needs to eat, whether or not it will harm a person, and what antibiotic will be effective against it. So, DNA sequencing provides the most precise knowledge possible to identify, classify, and generally understand a microbe.

At this time, researchers have deciphered a substantial portion of the "book of life" for many organisms but some aspects of the text are still unclear. While the revolution in the use of WGS to understand biology has greatly advanced our knowledge of microbiology there are still gaps, even in *E. coli*, the best studied microbe. Despite these gaps in knowledge, WGS of microbes will be the method of choice in microbial analysis for the foreseeable future. Even if we cannot perfectly understand the language of the whole book

[4] Flagellum (flagella – plural) is a hair-like structure that protrudes from a bacteria normally used for locomotion.

we can still accurately compare the books letter by letter. Therefore, WGS will likely be the gold standard method for characterising and comparing microbes.

3.2 What is NGS

Next Generation DNA Sequencing is a technology to rapidly read DNA sequences from many samples at once. It is the technology that underpins Whole Genome Sequencing – WGS. Development of this technology continues to improve and increase its output rapidly. The nature and scale of the data generated by NGS necessitates sophisticated computer analysis but this will be largely hidden from end users due to the development of automated data processing systems.

Next generation DNA sequencing, NGS, is a commonly used "catch-all" phrase to describe what could more accurately be described as the second generation of DNA sequencing technologies. It is not within the scope of this document to present a lengthy history of DNA sequencing technology but a useful review is available elsewhere (Shendure *et al.* 2017). Essentially, the move from first generation DNA sequencing technology to second generation was a transition from performing a single reaction for every sequence generated to a massively parallel system that allowed millions of DNA sequencing reactions to be done at once. Using a first generation DNA sequencing technology a DNA sample of interest was placed into a tube, a biochemical reaction was done and using some technology to visualise the results perhaps 1000 bases of DNA sequence data could be generated. In addition, just getting the DNA samples was often an onerous task requiring a range of genetic manipulations and purification steps. Using NGS technology a single laboratory benchtop sequencing machine can take a relatively crude tube of DNA and perform the equivalent of hundreds of millions of first generation DNA sequencing reactions. Another illustration of the magnitude of the change in scale of DNA sequencing technology is to compare the $2.7 billion (FY 1991 dollars) and approximately 12 years it took to sequence the 3 billion bases for the original human genome project compared to the performance of current generation instruments that can sequence 10 human genomes at once for less than $1000 per genome (note, microbe genomes are 1000 times smaller than humans). The advances leading to the current state of the art NGS technologies are the culmination of advanced engineering, biochemistry, and computational sciences. They were largely driven by the quest to rapidly and cheaply sequence the human genome which has led to advancements in excess of Moore's "law"[5] (the rate of DNA sequence data production is more than doubling every year).

NGS data creates significant computational challenges due to the abundance of data it creates as well as the analysis problems caused by breaking microbial genomes down into millions of small fragments. Whether attempting to reassemble the genome or simply comparing the data to previously known genomes these tasks are computationally intensive often requiring processing power far in excess of what standard desktop computers are capable of. In addition, both raw and processed data files are large requiring substantial resources be committed to data storage and archiving. A subspecialty of computer sciences called bioinformatics attempts to link computer programming, statistics, and biology to cope with data such as that generated by NGS and WGS. A wide range of processing systems have been developed and automated processing "pipelines" now exist that can take in raw NGS data and output human readable analysis.

[5] Moore's "law" is an observation not an actual physical law or rule that was coined during the rapid advancements in microprocessors and refers to the number of transistors in an integrated circuit doubling approximately every two years

3.3 Overview of NGS technologies

3.3.1 Short read NGS

Short read NGS sequencers generate a lot of data from many samples very inexpensively and will constitute the primary instrument type used by regulators and testing labs for the near-term future. As the name implies, this technology reads the DNA using a multitude of short segments (also called reads). The short segments it generates usually cannot be reassembled into a perfect and fully intact copy of the original but it provides abundant and accurate data.

The short read NGS technologies are typified by instruments produced by Illumina (e.g., HiSeq, MiSeq, NextSeq) and Thermo Fisher (e.g., Ion PGM, Ion GeneStudio). While the chemistry utilised by these instruments varies substantially between suppliers the underlying principle is similar. When supplied with fragmented and properly prepared pieces of DNA these systems are capable of generating millions of relatively short and accurate DNA sequences. Very little work is required to yield vast amounts of DNA sequence data. Although the first iteration of this technology yielded DNA sequencing reads of only 35 bases in length the current generation is substantially longer ranging from 150 bases to 400 bases. These reads cannot be readily assembled into a finished copy of the genome from which they were derived but a close approximation is possible. The length of these reads makes perfect assembly technically impossible for most genomes and the chemistries provide uneven coverage so some areas of genome may have limited data again limiting the ability to generate a finished genome. Despite these problems the NGS data is accurate and data rich. Despite the name, whole genome sequencing, most applications do not require a completely finished genome sequence. There are currently well-developed methods that permit multiple independent samples to be mixed on a single run of these instruments greatly increasing the number of samples that can be processed at one time. The short read NGS technologies are currently the method of choice for most WGS techniques. Most applications of NGS/WGS applied in the food sector utilize short read technologies with instruments from Illumina dominating the field at this time.

Advantages

- simple sample preparation
- simple sequencing reaction
- high throughput – many samples and large amounts of data per sample
- low sequencing error rate
- low cost per base

Disadvantages

- short reads have limited capacity to produce complete genomes assemblies

3.3.2 Single molecule real-time sequencing (SMRT)

The PacBio SMRT sequencers are capable of reading long strands of DNA accurately but at a much higher cost and for fewer samples than short read NGS. SMRT is primarily used for generating the reference sequences used for outbreak tracking and surveillance. This technology has the capacity to recreate a complete copy of the original DNA but is unable to process large numbers of samples.

Single molecule real-time sequencing (SMRT) used by PacBio instruments generates substantially longer DNA sequencing data than the previously mentioned technology. DNA sequencing reads averaging 15,000 bases are typical with lengths up to 100,000 bases possible. This technology has a somewhat higher level of sequencing error than the short read methods but the increased output in current iterations compensates

for this yielding similar level of accuracy to short read methods in the final data. The SMRT technology lacks the throughput needed for bulk testing and analysis of bacteria but is routinely employed to generate the reference data needed for isolate testing and comparison. This technology is likely to remain a reference and research tool rather than a mainstream testing technology.

Advantages

- reasonable throughput – low number of samples and modest to large amount of data
- modest sequencing error rate
- moderate cost per base
- long DNA sequencing reads permit complete or near complete reassembly of genomes

Disadvantages

- higher cost per base
- more demanding sample preparation
- more complex (difficult) sequencing reaction

3.3.3 Nanopore sequencing

Oxford Nanopore sequencers are capable of reading extremely long strands of DNA but do so very inaccurately and cannot deal with large numbers of samples. The instruments are very low in cost but the lower output of DNA sequence data yields a relatively high cost (per base) compared to short read NGS. Despite the deficiencies, low instrument cost and the simplicity of the Nanopore system may represent the predecessor to a generation of in-line / real-time DNA sequencers constantly searching for the presence of unwanted organisms.

Oxford Nanopore produce a range of instruments that directly read the sequence of DNA without the need for complex chemistry or enzymatic reactions like the above-mentioned technologies. These instruments are novel in both the method employed and the cost/form factor of the instruments. By employing a solid state technology approach the Nanopore instruments are about the size of a USB drive and are capable of reading the DNA sequence by directly examining the electrochemical properties of the individual bases as the DNA strand passes through the pore. This has the capacity to yield very long DNA sequencing reads (exceeding 800,000 bases for record runs). It would appear that sequencing of an entire bacterial genome in a single read is possible in the near term. The extremely long reads of Nanopore sequencing come at the cost of accuracy with error rates so high that one of the other sequencing methods is often used for error correction. In addition, Nanopore sequencing typically yields thousands of DNA sequencing reads compared to millions produced by short read NGS. This means that even with lower instrument cost the actual cost of sequencing is higher than the other methods.

Oxford Nanopore technology has been advancing rapidly but these instruments are currently limited to use for research purposes. Sample preparation is straightforward and can be automated with a small form factor microfluidic system. The instruments are inexpensive and can already be connected to smartphones and laptops for field use. It would seem highly likely that this technology or a successor to this technology will be employed as a real-time surveillance system. Potentially performing constant DNA sampling and sequencing producing a warning when unwanted sequences are detected. The first generation of such an apparatus (the Flongle[6]) has recently been released by Oxford Nanopore

Advantages

[6] https://nanoporetech.com/products/flongle

- long DNA sequencing reads permit complete or near complete reassembly of genomes
- simple operation and sample preparation (except for extreme read lengths)
- small instrument suitable for field, factory, or remote locations

Disadvantages

- higher cost per base
- high sequencing error rate
- low throughput – low number of samples and modest amount of data (per single unit)

4 What can NGS data be used for

In addition to the typing of microbial isolates and outbreak tracking, NGS-based methods can be applied to in wide array of ways. Some examples include:

1. ***In-depth characterisation of microbial isolates – a huge amount of information about a microbe is present in NGS data. Including information such as serotype data for typing, the presence of all genes that determine if a microbe is dangerous (and how dangerous it might be). The data can also be used to identify those microbes that are challenging to identify with traditional microbiological methods.***

2. ***Community analysis and metagenomics – which can determine what microbes are present in a sample without culturing. It is even possible to understand what capabilities members of the community have (e.g., resistance to antibiotics or disinfectants).***

3. ***RNA-Seq transcriptomics – determines how microbes react to the environment or some treatment. By knowing how microbes defend themselves provides a means to counter these defences.***

4. ***Accurate enumeration – accurately and specifically counting things that are otherwise difficult to count accurately. Even microbes that are nearly identical can be discriminated using NGS methods.***

5. ***Food adulteration / substitution – the genetics of food materials paints a very clear picture of what is or is not present in the final product.***

6. ***NGS generates the background data for all genetics – foundational data for human health, cattle genetics, feed genetics, and soil genetics to name a few.***

An increasing range of applications utilise NGS data across most areas that deal with living things in some manner from the clinical/veterinary fields, environmental, pharmaceuticals, food production, to agriculture. In addition to the WGS used for isolate tracking and outbreak analysis a selection of common applications is listed below. This is not an exhaustive list but represents applications currently in use with possible relevance to the food sector.

1. In-depth characterisation of microbial isolates
   Analysis of microbes using NGS technologies such as WGS yields a substantial amount of information. Whole genome sequencing data can now be used to determine the serotype of an organism replacing the tedious laboratory testing normally required. Another application of WGS data is the capacity to predict antimicrobial resistance from microbes. There are now publically available resources to predict antimicrobial resistance from WGS data. An important application of WGS data is the detection of genes involved in making pathogenic microbes virulent. By detecting virulence genes the degree of risk posed by an

organsim can potentially be estimated. It can be challenging to accurately determine virulence of a particular microbe in humans based solely on DNA sequencing data so these estimates need to be examined with some caution.

Another benefit of the abundance of data produced by WGS is the capacity to more accurately speciate an isolate. Many organisms are diffuclt to discriminate from closely related species using traditional laboratory methods and even single gene DNA sequencing methods[7] can yield ambiguous results. For example, *Clostridium sporogenes* and *Clostridium botulinum* are often problematic to discriminate but it is vitally important to accurately differentiate them. While traditional laboratory methods struggle to separate the two species, WGS data can readily separate them.

2. Community analysis and metagenomics

   The ability of NGS technology to produce abundant data and handle large numbers of samples make this a powerful tool for environmental and community analysis. Rather than focusing on a single isolated microbe, the composition of an entire community of organisms is determined. This is accomplished without the need to culture or grow the community. Instead, a specific region of every cell is amplified and these amplification products are then sequenced. The region targeted in bacteria is the 16S or small subunit ribosomal RNA. This region is useful because it is present in every known bacteria and regardless of how different the bacteria the 16S ribosomal RNA can still be detected and amplified. Although this gene is consistent across all the bacterial species, there is sufficient variation to discriminate at the level of species or genera. Hundreds of samples can be analysed for the presence of hundreds or thousands of different microbes. This methodology is at the forefront of both environmental and human health research (e.g., human microbiome analysis). Companies like Neogen are now offering this community analysis as a commercial service. It has been applied in AMPC funded research to understand the sources of carcass contamination at slaughter.

   A more in-depth version of this analysis that focuses on the capabilities of the community as well as their composition is called shotgun metagenomics. In this methodology, every microbe is sequenced in its entirety rather than just a characteristic region. Then every enzyme and every protein that can be produced is known, allowing analysis of what biochemical reactions the community is capable of. By employing this procedure, one could examine a broad range to things like the nature of antibiotic resistance residing in the community or the potential of a community to produce toxins. Although shotgun metagenomics provides unprecedented levels of information about a community it lacks the sensitivity of the more targeted amplification approach and comes at a much higher cost and for fewer samples. The sensitivity is limited in shotgun metagenomics because low abundance microbes are overwhelmed by abundant organisms and will not be detected unless extraordinary quantities of DNA are sequenced.

3. RNA-Seq or transcriptomics

   This technique uses NGS technology to examine the way that microbes react to their

[7] The single gene DNA sequencing referred to here is 16S rRNA sequencing. This is a test done at specialty laboratories using older generation DNA sequencing instruments that run single DNA sequences rather than the massive numbers that NGS instruments do. The 16S rRNA gene is often used to determine the species of a microbe.

environment or other stimuli. Rather than sequence the DNA of a cell, the RNA[8] which is produced to make proteins is instead sequenced. For example, when a microbe is treated with disinfectant it attempts to activate systems to protect itself. Usually this entails genetic signals that are expressed as RNA that is then used by the cell for producing a range of proteins in response to the disinfectant. RNA-Seq detects the signals during this change in expression of proteins so we can understand how the cells are attempting to overcome the disinfectant and take steps to counter the cells defences.

4. Accurate enumeration
   The output of NGS technology quantitatively represents the input material so it is a powerful and accurate tool for counting things (that contain DNA or RNA) in the life sciences. A large amount of data is produced that is directly representative of DNA provided to the reaction meaning clear statistically supported numbers can be determined. For example, NGS can be used to count how many *E. coli* are in a broth containing a mixture of other species more accurately than quantitative polymerase chain reaction (PCR) but with less sensitivity. Quantitative PCR can detect a smaller number of cells than direct counting of DNA can due to the amplification inherent to the PCR process).
5. Food adulteration / substitution
   NGS data can be used to unambiguously identify the living material in many food products, particularly those subject to substitution or adulteration with lower value products. In addition it can be used to detect the microbial community associated with many foods (e.g., the microbial community from a particular farm could be discriminated from another)
6. NGS generates the background data for all genetics
   NGS technology permits the WGS to be determined for both bacteria as well as higher organism. This WGS data is then the foundation for human health, cattle genetics, feed/pasture genetics, soil community genetics, and plant/crop genetics. In addition testing systems like the Neogen Neoseek STEC testing system which uses a mass spectrometer to analyse isolates is built upon WGS data.

[8] RNA is the molecule produced from the DNA of all living things to encode proteins. The genome contains the genes that make the proteins. When the cell wants to respond to a stimulus it triggers the production of RNA spanning the desired gene of interest. This RNA is then converted into protein by the normal cellular mechanisms.

5 NGS/WGS for surveillance, outbreaks, and source tracking

Whole genome sequencing data is already in use around the world. The most mature WGS based system is the GenomeTrakr network in the US that has already been instrumental in regulatory interventions against food processors. In addition to Australian microbial isolates tested in the US, several Australian institutions contribute WGS data to the publically available GenomeTrakr network. There are several different ways to analyse WGS data but all rely upon the basic concept that closely related microbes have very few differences in the WGS data while unrelated ones have many more.

The use of NGS technology and WGS in food microbiology and regulation is now commonplace in North American and European nations but is not limited to these countries. The most mature example of the application of WGS to food regulation is the approach taken by the U.S. Food and Drug Administration (FDA). FDA's Food Safety Inspection Service use WGS to sequence a growing list of isolates from selected pathogen species (e.g., *Listeria, Salmonella*) as well as all isolates categorised as adulterants. The FDA's GenomeTrakr Network now provides vital information for epidemiological investigations and assists in tracking the source of microbes isolated from food to within relatively small geographic areas or potentially to the individual food processing facilities. The FDA have acted on several occasions to close processing facilities based largely on WGS evidence and the frequency of such events will undoubtedly increase. A core goal of the GenomeTrakr Network is the pairing of WGS data from foodborne pathogens to geographic source data for the isolate. It is hoped that this information will permit more rapid responses to outbreaks such that (as stated by the FDA), *"The faster public health officials can identify the source of contamination, the faster the harmful ingredient can be removed from the food supply and the more illnesses and deaths that can be averted."*[9]. Although the database of microbial isolates subjected to WGS is large and growing rapidly there is some concern about the uneven sources from which the samples are derived. The FDA states, *"The need for increased number of well characterized environmental (food, water, facility, etc.) sequences may outweigh the need for extensive clinical isolates"*[10].

While the GenomeTrakr Network is largely driven by US based laboratories and resources there are a substantial number of non-US affiliates, three of which are in Australia. These include the Doherty Institute in Melbourne along with University of New South Wales and the Institute of Clinical Pathology and Medical Research at Westmead Hospital. The sequence and associated analysis data entered into the GenomeTrakr Network is available to the general public.

Outbreak analysis and the comparison of microbial isolates using WGS can be accomplished by a number of different methods such as Single Nucleotide Polymorphism / Single Nucleotide Variant (SNP/SNV), Whole Genome MLST (wgMLST), and Kmer analysis. The technical details of these methodologies is not crucial to this briefing. All of the methods begin with the same WGS data derived from an isolate and depend upon the premise that two isolates from the same source should be very similar if not identical. An isolate from a different source is more likely to be different. The nature of these differences are derived from the natural biological forces that drive evolution and adaptation in microbes. As time passes growing microbes tend to accumulate small numbers of changes to their genome sequence that are passed on to progeny. Also, microbes that live in different isolated locations follow a different paths as they evolve over time again leading to differences in the genome sequence. A significant problem with using PFGE as the standard

[9] https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/default.htm

[10] https://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm403550.htm

method to discriminate between closely related strains was that in many cases it lacked the capacity to discriminate isolates in an outbreak from those not involved in the outbreak. This problem was particularly acute in *Salmonella* where PFGE was often ineffective in discriminating between outbreak organisms and unrelated organisms. A highly simplified description of how WGS data can be used to discriminate between isolates when tracking the source of an outbreak is described below.

> *WGS data from three Salmonella isolates are examined. A is from a processing facility while B and C are from food samples. NGS data says A and B only differ by two changes in their genomes while C differs by 1*
>
> *50 changes from A and B. Therefore, this suggests that isolate A from the processing facility is closely related to food isolate B while food isolate C is unrelated to either A or B.*

As stated in the example above, the scenario presented is simplified and masks the subtlety involved in deciding what is and what isn't related. That decision is impacted by the organism being examined as well numerous other factors. It must be remembered that NGS data, like PFGE data, cannot be used as the sole source of evidence in determining the origin of an isolate. Outbreak investigations remain dependent upon epidemiological analysis, NGS simply provides a more accurate and rapid set of tools to facilitate the investigation.

5.1 The Canadian example for implementation of NGS/WGS

> ***Canada has a well-developed system using WGS data for food safety microbiology. The modular WGS based system implemented at the Canadian Food Inspection Agency reduces the time and cost of the analysis of microbial isolates while increasing the amount of data that is generated.***

Canada is a useful example of the successful implementation of NGS/WGS into the food safety system for a country the size of Australia. Canada has a similar population size (approximately 1/3 larger), an English colonial history, and an economy with a large dependence on primary production. One substantial difference is the existence of national bodies like the Canadian Food Inspection Agency (CFIA) and Genomics Canada for which there is no equivalent in Australia. The CFIA is responsible for food safety compliance and outbreak investigation with six large CFIA labs as well as a range of smaller provincial facilities.

NGS implementation has largely replaced PFGE and serotyping. Isolates are no longer transported between labs, only the WGS data is moved electronically. GeneSippr is a high speed analysis system that can sample NGS data while the WGS is being determined, to essentially perform well defined laboratory analyses that would have previously been done by polymerase chain reaction (PCR). Using in-built quality control standards it simply replicates a laboratory test in the computer using the NGS data. Upon completion of the WGS analysis, data is transferred to the GeneSeekr program that identifies virulence markers, performs computer based serotyping, performs molecular typing (using both MLST and K-mer based approach), searches for antimicrobial resistance markers, and presents a genomic risk assessment report.

Implementation of GeneSippr and GeneSeekr has led to cost and time savings for the CFIA along with the benefits of using a more accurate method of outbreak investigation. Lacking an overarching government food safety body and in the absence of government financial support it is likely that Australian state labs will have to develop similar systems either in collaboration with other countries or independently.

6 Issues caused by NGS and WGS

- ***The need to devise new regulations and standards (for technology, methods, analysis, and data storage).***
- ***The requirement for development of new testing regimes and naming standards to assure consistency of results between testing facilities.***
- ***Allocate the resources required to store large amounts of WGS and computer analysis data (in the cloud or locally).***
- ***Fulfil the need for training of testing lab staff, quality control staff, and regulators to cope with new data types.***
- ***Decisions concerning who will have access to WGS data (public vs. private) will be required.***

The application of NGS and WGS to the analysis of foodborne organisms creates of number of challenges that must be addressed.

- The application of WGS to foodborne organism analysis necessitates the development of a host of new standards and regulations. Key areas of interest include:
    - The development of globally harmonized regulations and guidelines for the collection and analysis of WGS data will be required. This is likely to be a challenging task given that several methods of analysing WGS data are already in use. For example, the PulseNet network is currently implementing whole genome MLST while FDA's GenomeTrakr network uses SNP.
    - The laboratory and data analysis methods will need to be validated to appropriate international standards. This represents a formidable challenge since much of the software has developed in an ad hoc fashion in a rapidly evolving field. Considerable effort will be required to validate the software and the sophisticated algorithms used for WGS analysis. It will be challenging to ensure consistency in the collection and analysis of data given the variety of NGS methods available as well as the large selection of analysis tools.
    - Standards on the storage and availability of WGS data will also have to be developed. This may be a contentious issue with regulators likely to encourage full public access to this data while industry bodies are likely to be uncomfortable with that level of disclosure and the possible legal ramifications associated with it. In addition, given the extensive use of cloud-based storage systems there is the potential for challenges to arise due to the storage of data in locations outside the regulatory controls of the locality in which the data was generated.
    - It is likely that there will be issues around the access to non-government / regulator generated WGS data. Would a producer be required to give WGS data they held to a regulator? Would a regulator use producer generated WGS data as part of an investigation? While these matters will need to be resolved it should be

remembered that WGS data is simply another typing method that the industry will adapt to.

- The advancement of WGS has led to a decline in the need for traditional antibody based serotyping which has been a mainstay of microbiology. Serotyping is a time consuming and expensive process that is an important early step in classifying an isolate. A WGS based replacement for serotyping will need to be validated and linked to the previous system. An entirely new naming scheme is likely to be needed since organisms like *Salmonella* and *E. coli* are generally named based on their serotyping data. Once serotyping is fully replaced by WGS data new "types" will likely need new agreed naming system. Work in this area is well underway for *E. coli (Joensen et al. 2015)* and *Salmonella (Yachison et al. 2017)*. Whole genome MLST is a likely contender for a replacement nomenclature for serotyping but it is premature to predict a final outcome at this time.
- NGS/WGS generates large amounts of data from both the raw data generated by the sequencing instruments and the processed data. Raw instrument data files are extremely large approaching the terabyte range and are not generally maintained by end users. Files containing the NGS reads are the form of data that is normally stored and these files are typically hundreds of megabytes to the gigabyte range. This file size is compounded during the data analysis steps as filtering, pairing, and other steps result in modified versions of the data file that must all be maintained for quality control purposes. It is likely that this data will have to be maintained for extended time periods so provisions will have to be made for the archiving of this information. This will require decisions to be made about where the data is maintained (physical local hard drives or cloud based storage media). If cloud based storage is used this raises some concerns about where this data actually resides and who has access or control over the data. In addition to storage, some provision will have to be made for the computational data infrastructure. Again, the required computer resources could be either local or provided by a cloud based service.
- It is likely that some retraining will be required for dealing the NGS/WGS data. Although WGS data can be readily processed into human readable forms adequate training will be required to interpret the output from analysis. The laboratory based methods are substantively different from those required for routine microbial testing. Regulators and quality control staff will need to adapt to interpreting novel types of information as well.
- PFGE data used by the global PulseNet network has been the "gold standard" for surveillance of foodborne organisms. A transition from PFGE to a more informative WGS based method is currently being investigated. One impact of this will be that there will be a push to make the more informative WGS data publicly available in the same way that PFGE patterns are made available. Data access will be a significant concern as the WGS is increasingly used in the public and private sector.
- The digital nature of WGS data raises issues around matters such as privacy and how it may be used in otherwise unexpected ways. A list of possible scenarios of interest suggested by the FAO (adapted from the FAO report on the applications of WGS in food safety management)[11] are listed below.
    1. Legal issues

       Although there is a sound scientific basis for the accuracy and reliability of WGS data it

[11] Applications of Whole Genome Sequencing in food safety management http://www.fao.org/3/a-i5619e.pdf

has not been fully codified in regulatory frameworks. The diversity of methods (both laboratory and computational) have not been harmonized so there may be room for legal issues surrounding this area.

2. Trade issues
   At this time there are no international standards for either the laboratory or data analysis methods in WGS. Various international organisations are at work on this but the rapidly evolving technology and analytical methods makes harmonization extremely challenging. The lack of agreed methods and standards leaves room for this technology to be used (or abused) to create trade issues.
3. Proficiency testing
   The lack of harmonized methods described above makes quality validation and certification of WGS analysis difficult. A number of international groups are attempting to propose guidelines on how this could be accomplished.
4. Training and education
   Regulators, quality control personnel, testing agencies will require training in the interpretation, analysis, and management of WGS data.
5. Communication issues
   All stakeholders from government to industry to the general public will need to understand the utility and potential consequences of the application of NGS technologies to routine testing and surveillance.
6. Continuous improvement
   NGS technology is still rapidly improving so it will require regular review to ensure that improvements are incorporated into the approved methods and analysis pipelines.

6.1 Local issues for the Australian Beef Industry

- ***In Australia there is a minimal amount of WGS data for isolated microbes from food, processing facilities, and the environment. This in turn limits our knowledge of how diverse these microbes are which means it will be difficult to know if a microbe found by one producer is unique to a single facility (region, state) or if it is found everywhere. This may increase the risk of misattribution in the case of a foodborne illness incident.***

- ***WGS data may permit Australia to capitalise on the limited risk posed by local pathogenic microbes (when compared to the same organisms from overseas).***

- ***Australian regulators and health labs have begun adoption of NGS/WGS later than many parts of the world and this may limit the role Australia can play in how this technology is implemented in international standards and regulations.***

6.1.1 Limited baseline WGS data

There is limited WGS baseline data for Australian isolates compared to the situation in North America and Europe where some countries now have many years-worth of data. Baseline data includes isolates from food, illnesses, and the environment. Without this information the level of diversity (amount of difference) between environmental, foodborne, and illness related isolates is unknown. Diversity or the differences between isolates is what NGS methods like WGS measure. As stated previously, the US

FDA's GenomeTrakr network lacks sufficient environmental samples to understand the amount of diversity in the environment. At this time, Australian WGS data is limited to some state laboratories and research organisations such as universities and the CSIRO. Australia has no national body or system for the adoption of NGS technologies as part of standard pathogen testing. While there is NGS activity in selected state labs, the course this technology will follow in Australia remains uncertain. There is a high level of interest in WGS but it is likely that adoption of NGS and routine WGS of microbial samples by state health labs will initially be limited to health and outbreak related samples. It is unlikely that environmental and routine samples will be part of any database or tracking system. There is the possibility of some trade or market access related issues arising for Australia and Australian industries using lower resolution testing methods than those applied by countries that are key markets.

6.1.2 Reduced virulence of some Australian pathogens

Current research suggests that indigenous pathogens such as Australian enterohaemorrhagic *E. coli* are less pathogenic than those found elsewhere. In addition, Australian *Salmonella* tend to have lower levels of resistance to medically important antibiotics. While these lower pathogenicity, lower risk organisms would still be classified as pathogenic, the likelihood of them causing severe disease should be reduced. Previous lower sensitivity testing and typing methods cannot adequately distinguish Australian isolates from those derived from elsewhere. WGS based methods can readily distinguish the less pathogenic Australian enterohaemorrhagic *E. coli* from overseas isolates. There may be some tangible benefits to Australian producers to have a body of evidence demonstrating the prevalence of what could be described as a lower risk variant of a pathogen.

6.1.3 Australia's late adoption of NGS data

Public health labs are the force driving NGS uptake in Australia and seem likely to lead the implementation of this technology. While NGS technology and the utility of WGS for microbial analysis has been recognised, Australia lags behind many other countries including important trade partners like the US. The GenomeTrakr network started by the FDA in the US has been operating since 2013 and now contains WGS data for hundreds of thousands of isolates. In Australia the public health sector is fragmented by state with no organizing national bodies like those in North America and Europe. This sector is only now taking its first tentative steps toward adoption of NGS/WGS. This late adoption may have knock-on implications for Australia's role in determining international standards and policies around the use of NGS/WGS data.

7 Benefits of WGS

- ***WGS is superior to all previous methods for the tracking and surveillance of microbial isolates.***
- ***The high sensitivity of WGS methods may permit the discrimination of local variations in microbes so it may be possible to tell if a bacterium is from producer A or producer B. Although theoretically possible it requires more evidence than simply WGS data from two bacteria to establish this as fact.***
- ***Implementation of WGS methods for the analysis of isolates should make analysis both quicker and cheaper.***
- ***The information provided by WGS makes the detection of emerging microbiological threats easier since you do not have to know what you are looking for to detect something new.***
- ***WGS provides a wealth of data about an isolate making some additional testing redundant. Automated analysis packages now exist that can provide this information in a human readable form to end users.***

- WGS is now the method of choice for the tracking and surveillance of outbreaks. The analysis of bacterial isolates using WGS based methods have been proven to be superior to the previous standard established with by PFGE (Moura *et al.* 2017). A higher level of certainty on the source tracking/outbreak isolate tracking is due to the greater amount of the genome examined by WGS.
- The high level of sensitivity in WGS based typing makes it possible in some organisms for WGS data to detect differences between the microbial populations at different localities. While the capacity to use WGS data to determine where an organism is fromis a key goal for the US GenomeTrakr system caution must be exercised. The mechanisms of transport/movement of microbes around the environment is highly variable with both natural and human forces playing a role. WGS data needs to be considered in the context of standard epidemiological analysis to avoid possible misattribution of organisms.
- More rapid and less expensive identification and tracking of outbreak isolates is possible using WGS based methods. It is possible that there will be further increases in speed as sequencing without isolation increases. By eliminating the need to first culture an isolate, outbreak analysis could theoretically be reduced to hours (this remains theoretical at this time).
- WGS based methods are capable of detecting emerging food-borne pathogens. Conventional testing methods, particularly the molecular methods such as PCR only search for known targets but WGS does not rely upon previously known information.
- WGS analysis provides far more data than previous methods. In a public health setting the WGS data for an isolate is generally entered into an automated analysis system that provides a wealth of data. There are a number of systems in use around the world. Large public health and clinical labs in Australia utilise an Australian software package (Nullarbor – https://github.com/tseemann/nullarbor) to determine the following from WGS data:

- quality of DNA sequence data
- identification of the species
- typing by relevant typing scheme (e.g., MLST)
- presence of antibiotic resistance genes
- high resolution genome comparison (e.g., SNP phylogeny)
- detailed analysis of both common and novel genes present

8 Recommendations

Next generation DNA sequencing technologies such as WGS do not pose a significant threat to standard practice of the red meat industry but impact on regulations and testing regimes should be followed closely.

The public health sector is driving the implementation of NGS/WGS technology in Australia, industry should advocate to have input on the development of new regulations and policies that will undoubtedly affect them.

Industry should advocate for changes to the required testing methods to capitalise on the likely changes (such as cost savings) that may come with the uptake of NGS/WGS technologies.

Government agencies such as FSANZ should be queried to assure Australian interests are being supported as global entities develop new standards and regulations to deal with NGS/WGS implementation.

Whole genome DNA sequencing data has the potential to provide tangible evidence for the reduced risk posed by some pathogenic microbes found in Australia. This evidence could be the basis for lobbying to reduce the stringency of testing required for meat from Australia.

- Next generation DNA sequencing technology and WGS are analysis and typing tools that are improvements on existing methods. These new technologies and methods do not represent a threat to existing business practices for the red meat industry and the drastically improved accuracy should result in a reduction in false positive attributions. Although the increased sensitivity of WGS methods can link historical isolates to outbreaks they still require solid epidemiological data just like the current methods. Even if WGS data shows that a problematic isolate matches an organism found in a production facility this on its own is insufficient evidence to attribute “blame”. There are numerous scenarios that can lead to such a match (e.g., the organism lacks diversity and the target strain is widespread). While not disruptive to the red meat industry practices NGS technology is likely to have significant implications for regulators, government, and testing laboratories (public and private). NGS/WGS will require these bodies to substantially change practice, retrain the workforce, and alter regulations and standards (this will in turn impact upon industry). Therefore, it would be prudent for the red meat industry to be aware of the regulatory and methodological changes that are going to indirectly impact their industry.
- Public health labs are the force driving NGS uptake in Australia and seem likely to lead the implementation of this technology. In Australia this public health sector is fragmented by state with no organizing national bodies like those in North America and European countries. The public health sectors priorities and requirements are likely to differ from that of the red meat industry. Although the public health sector would like industry involvement they will likely proceed regardless of the level of industry involvement. It is recommended that the red meat industry advocate for clarity on the matters such as those listed below.

    i. Who will have access to NGS data (and associated descriptive data aka metadata)?
    ii. Will WGS data from certain microbes require public release?
    iii. How regulations will adapt to culture independent methods (no isolate(s) will exist)?
    iv. What will be required of industry and testing labs in a post-NGS testing regime?
- NGS technology is likely to generate cost and time savings in the near future. The cost for NGS is in a downward trend and the labour requirements are minimal compared to traditional microbiological testing methods. Industry should advocate for changes to the testing regime to capitalise on these potential cost savings.
- Government (e.g., FSANZ) and trade bodies should be queried to determine if Australia has a seat at the table during the process of developing new regulations and standards to deal with the increasing application of NGS/WGS data. There is the potential that WGS data could be used for the benefit of Australian trade by providing strong evidence of the safety of Australian products. Alternatively, Australia's slow uptake of NGS applications could potentially be used as a potential trade barrier.
- The increased information about the pathogenic potential of microbes that is gained from WGS analysis has potential to provide the required evidence for a more nuanced treatment of organisms. No longer is an isolate just a particular species or serotype, now the array of antibiotic resistances, and other virulence markers will be fully revealed. Given current data, Australia would be well placed to advocate for a number of "local" pathogens posing significantly lower risk than overseas derived isolates. An example of this would be the case for Australian enterohaemorrhagic *E. coli* being less virulent than those found elsewhere in the world. As NGS adoption increases, the red meat industry should push for an evidence based determination of the risk posed by those organism for which testing is required.

9 Useful resources for additional information

1. Applications of Whole Genome Sequencing in food safety management. Food and Agriculture Organization of the United Nations in collaboration with the World Health Organization http://www.fao.org/3/a-i5619e.pdf
2. Food Safety and Inspection Service FY 2018 Annual Plan https://www.fsis.usda.gov/wps/wcm/connect/88c7a056-3f26-41e1-8f03-d18bc9e0fb05/Annual-Plan-FY2018.pdf?MOD=AJPERES
3. The future of DNA sequencing. Eric D. Green, Edward M. Rubin, Maynard V. Olson. Nature 550, 179–181 (12 October 2017) doi:10.1038/550179a https://www.nature.com/news/the-future-of-dna-sequencing-1.22787?WT.ec_id=NATURE-20171012&spMailingID=55119582&spUserID=MjA1NjE2ODM2OQS2&spJobID=1261947732&spReportId=MTI2MTk0NzczMgS2
4. EFSA Scientific Colloquium N°20: Whole Genome Sequencing of food-borne pathogens for public health protection https://www.efsa.europa.eu/en/events/event/140616

10 References

Joensen, KG, Tetzschner, AMM, Iguchi, A, Aarestrup, FM, Scheutz, F (2015) Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *Journal of Clinical Microbiology* **53**, 2410-2426.

Moura, A, Tourdjman, M, Leclercq, A, Hamelin, E, Laurent, E, Fredriksen, N, Van Cauteren, D, Bracq-Dieye, H, Thouvenot, P, Vales, G, Tessaud-Rita, N, Maury, MM, Alexandru, A, Criscuolo, A, Quevillon, E, Donguy, MP, Enouf, V, de Valk, H, Brisse, S, Lecuit, M (2017) Real-Time Whole-Genome Sequencing for Surveillance of Listeria monocytogenes, France. *Emerg Infect Dis* **23**,

Nadon, C, Van Walle, I, Gerner-Smidt, P, Campos, J, Chinen, I, Concepcion-Acevedo, J, Gilpin, B, Smith, AM, Man Kam, K, Perez, E, Trees, E, Kubota, K, Takkinen, J, Nielsen, EM, Carleton, H, Panel, F-NE (2017) PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill* **22**,

Shendure, J, Balasubramanian, S, Church, GM, Gilbert, W, Rogers, J, Schloss, JA, Waterston, RH (2017) DNA sequencing at 40: past, present and future. *Nature* **550**, 345-353.

Yachison, CA, Yoshida, C, Robertson, J, Nash, JHE, Kruczkiewicz, P, Taboada, EN, Walker, M, Reimer, A, Christianson, S, Nichani, A, Nadon, C, Comm, PCS (2017) The Validation and Implications of Using Whole Genome Sequencing as a Replacement for Traditional Serotyping for a National Salmonella Reference Laboratory. *Front Microbiol* **8**,

Part 2: Implementation of Next Generation Sequencing in Food Microbiology Testing in Australia

November 2017

Dr Amy Jennison and Helen Smith

Public Health Microbiology

Forensic and Scientific Services

Queensland Department of Health

11 Tables

Table 1. Processes around sequencing data and minimal metadata stratified by entity type.

Table 2. Future directions for whole genome sequencing on microbial food isolates stratified by entity type.

12 Glossary

WGS	Whole Genome Sequencing
NGS	Next Generation Sequencing – technologies used to generate whole genome sequence data
wgMLST	whole genome MLST
wgSNP	whole genome single nucleotide polymorphism
cgMLST	core genome MLST
PFGE	Pulsed Field Gel Electrophoresis
MLST	Multi Locus Sequence Type
NATA	National Association of Testing Authorities
PT	Proficiency Testing

13 Summary

This report describes collated data from a phone survey conducted with major Australian laboratory and stakeholders involved in testing of microbial isolates from food products (meat industry) around implementation and use of whole genome sequencing (WGS) for bacterial testing and typing.

The survey focused on providing a snapshot of the current and future state of WGS based testing in Australian public health laboratories, commercial food testing laboratories and government regulators.

While three of the five Public Health laboratories are utilising in-house genomic sequencing, none of the 24 commercial testing laboratories involved in testing for the meat industry currently have any capabilities or any current plans to implement such capabilities.

Regulators and policy makers are in some capacities already utilising genomic sequencing data of microbial food isolates, particularly around outbreak investigations, and are keen that industry involvement in new technology does occur and that the new technologies are seen as a benefit and positive step forward for industry.

1 Background

Whole Genome Sequencing (WGS) technology advances are precipitating changes to the microbial regulatory and testing landscape globally. The use of whole genome sequencing approaches such as wgMLST and wgSNP analysis are replacing traditional methods like Pulsed Field Gel Electrophoresis (PFGE) for the typing of isolates. In addition, these methods provide unprecedented levels of accuracy in tracking the sources of outbreaks. When combined with new culture-independent techniques the implementation of Next Generation Sequencing (NGS) based methods is transforming microbial testing.

Stakeholders in the red meat industry have heard about the potential changes coming to microbial testing but lack a clear understanding of where the field is heading or what the ramifications of the new technology are. There are concerns about the greater powers to attribute microbial isolates to particular facilities and how this technology might impact upon regulations in the future. The red meat industry, through Meat and Livestock Australia (MLA) has decided to embrace the technology proactively.

The Meat and Livestock Australia (MLA) funded research project run by CSIRO, of which this survey belongs to, is designed to help the red meat industry understand the changes that WGS technology is bringing to microbial testing. By providing a clear picture of where WGS technology is headed in the microbial testing field, it is hoped that the red meat industry stakeholders will have a greater knowledge and understanding of the benefits, risks and limitations of NGS driven testing regimes.

2 Objective

To provide a snapshot of current and future state of WGS based testing of food borne pathogens across Australian public health state reference laboratories, commercial food testing laboratories to the meat industry and government policy makers and regulators.

3 Survey Method

Contacts were approached initially by email or phone explaining the project and asking for an interview answering around eight broad questions around their current use or application of WGS technology and/or data, and how this pertains to food isolates and raw product, with a focus on Shiga Toxigenic *E. coli* (STEC) and Salmonella.

Contactees were asked if they were not the appropriate person in their organisation to be the interviewee if the appropriate person could be suggested. Each organisation was contacted a minimum of twice.

Participants were assured that while persons/organisations contacted would be listed in the report that all answers will be stratified by organisation type (eg. Private laboratory, legislators, government laboratory) and participants would be non-identifiable by individual response.

Organisations contacted who were not currently performing WGS, and had no plans to do so within the next three years, were not asked to complete the full interview unless they wished to give further information.

A telephone interview was then conducted utilising the questions in *Appendix 2* and the results were recorded. Questions were developed in consultation with the CSIRO Industry Strategy project team.

4 Survey Results

4.1 Current State

24 Australian commercial microbial food testing laboratories with meat industry clients were contacted. 24/24 (100%) indicated that they were not currently using whole genome sequencing for testing and did not have any immediate plans to implement it. General consensus was that their clients were not requesting this type of work and they did not feel this work was within the scope of the “routine” laboratory but rather a higher level tool.

Five Australian Public Health Microbiology state reference laboratories were contacted. 3/5 (60%) laboratories possessed in-house whole genome sequencing capabilities and 2/5 have no in-house sequencing equipment and have no current plans to purchase. All laboratories have performed sequencing on microbial food isolates. 2 of the 3 Public Health Microbiology laboratories performing WGS were sequencing microbial food isolates only in response to disease outbreak investigations or research project work, while the third also receives referred food isolates for sequencing as part of surveillance testing.

Five Regulators involved in microbial food testing were contacted. All five (100%) are sending or requesting isolates to be whole genome sequenced (by research or Public Health laboratories depending on purpose) and using the results in some capacity, typically for either outbreak investigation or routine surveillance.

Two other entities were interviewed. One being a microbial food testing laboratory involved in government surveillance who had no in house capabilities but were referring isolates for testing at a Public Health Microbiology reference laboratory. The other being a surveillance related organisation who is not performing whole genome sequencing but is coordinating research projects around food related microbial isolates.

4.2 Data Access and Metadata association

Interviewees were asked about their use of metadata in association with genomic sequence. All respondents who perform whole genome sequencing or who receive sequence data maintained any identifiable metadata within their own secure computing environment. A tiered level approach was applied to metadata by all responding Public Health laboratories, with the most information being provided to government surveillance groups.

Any sequence made public was only associated with “minimal metadata” eg country/state, year of isolation and specimen type such as “environmental” or “non-human”. Respondents indicated that associating further levels of metadata with sequence in the public forum required governance review and/or discussion with industry.

The use of and policy around sequencing data is summarised in Table 1.

	Commercial laboratory	Public Health laboratory	Regulator	Other
Data stored on site	No	3/5 = Yes	No	No
Cloud used for data storage/analysis	No	2/5 = Yes	No	No
Access controls	NA	Staff only with tight controls	NA	NA
Period of time for data storage	NA	1/5 = 7 years 1/5 = 20 years+ 1/5 = indefinite	1/5 = 7 years 4/5 = indefinite	NA

Table 1. Processes around sequencing data and minimal metadata stratified by entity type.

4.3 Accreditation

Two of the five Public Health laboratories have recently achieved NATA accreditation for the generation of sequence data by whole genome sequencing. One other Public Health laboratory is preparing to apply for accreditation with 12 months. No other laboratories indicated intentions to pursue accreditation for WGS.

100% of survey respondents asked indicated that they considered the use of a NATA accredited laboratory for WGS based analysis of food microbial isolates as very important and acknowledged that quality control (QC) practices for WGS, while complex, are crucial for confidence in results produced by laboratories performing WGS.

All three of the Public Health laboratories participate in whole genome sequencing proficiency testing (PT) that encompasses food borne related microorganisms (Global Microbial Identifier PT)

4.4 Methodology

Interviewees were asked to name WGS based methodologies and analyses that they either had hands on experience with or were familiar with.

These included:

- SNP based typing used for phylogenetic cluster analysis
- core genome MLST (cgMLST) or whole genome MLST used for phylogenetic cluster analysis
- Bionumerics. A commercial software platform (http://www.applied-maths.com/bionumerics) used for genomic sequence analysis and typing.
- Ridom SeqSphere+: A commercial software platform (http://www.ridom.com/seqsphere/) used for sequence analysis and typing.
- Geneious: A commercial software platform (https://www.geneious.com/) used for sequence analysis.
- Nullarbor: an open source bioinformatics pipeline used to produce sequence analysis relevant to public health surveillance including quality control information, in silico derived MLST, de novo assemblies, SNP

based typing, core and accessory genome analysis (https://github.com/tseemann/nullarbor). This pipeline is in use in all three of the Public Health Microbiology reference laboratories performing WGS currently. This was the most common response from respondents.

4.5 Turn Around Time

All laboratories providing whole genome sequence data reported that their sequencing analysis is approaching real time. One regulator who requests genomic sequencing on food isolates is handling genomic data in near real time.

Other laboratories and regulators are only utilising sequence data in a retrospective fashion, not timely enough for intervention or follow up.

4.6 Future Directions

	Future Directions
Public Health Reference laboratories	• Using deep sequencing to address culture independent diagnostic testing • Performing metagenomic sequencing directly on food samples to look for bacteria present (useful for fastidious organisms) • Transition of all bacteria surveillance to WGS, including Salmonella
Regulators	• Feel that future directions will be drive by need such as determined by national and international standards and regulations • Desire for central database or publicly available sharing of sequence data from all laboratories and entities to assist in surveillance • Wish for openness on methods to assist less advanced laboratories in developing capabilities
Commercial laboratories	• May consider implementing WGS if there is an appetite from clients • Still consider WGS to be a research tool rather an application for industry

Table 2. Future directions for whole genome sequencing on microbial food isolates stratified by entity type.

4.7 Open Comments

- There is a need for epidemiologists who understand the technique and the analysis from WGS to ensure that the data is correctly interpreted.
- WGS will be a useful tool for providing confidence in outbreak source tracking. It has been shown that it in fact reduces the size of outbreaks through early detection. Therefore, outbreaks can be less severe and ended rapidly, reducing implications for food providers/processors. (see supporting references [1-3])

- Desire for standardisation, particularly amongst Public Health reference laboratories so that multi-state outbreaks or international surveillance can be performed.
- Cost needs to come down to make this technology viable for smaller laboratories.
- There is a need for industry to see the benefits of utilizing the technology within their own systems.
- Industry engagement is very important and more discussion is needed.

4.8 Themes identified during Interviews

- Public health laboratories are leading the way in the implementation of whole genome sequencing in both their capabilities around in house sequencing and in their accreditation status.
- Regulators have an appetite for the sequence data produced from WGS although they are only just beginning to address integrating this form of data into their surveillance. This appetite is in some ways being driven by the requirements of international governments and regulatory bodies.
- Commercial laboratories identified in this study as being involved in microbial food testing have no WGS capabilities and do not see the need to develop these capabilities in the short term future. Technology changes that drive cost reductions and ease of implementation are likely in the next five to ten years and may impact on commercial laboratory workflow. Most commercial laboratories were focused on current client requirements rather than potential future requirements.
- Metadata associated with sequence is being tightly controlled and only associated with sequence in house by public health laboratories and regulators. Some minimal metadata (typically country and/or state of origin and sample type eg “non-human”) may be associated with sequence for publication and this is being overseen typically by governance protocols in house.
- Regulators and policy makers were keen that industry involvement in new technology does occur and that the new technologies are seen as a benefit and positive step forward for industry. Very important that implementation not seen as a punitive but rather a focused response.
- Regulators and policy makers were keen that industry learnt of new technologies/techniques but were not sure that they (regulators and policy makers) were the right people to inform on the new technologies.
- Regulators and policy makers gave some instances such as the implementation of new technologies and collaborative work that has been done with egg producers particularly in South Australia as some of the potential positive outcomes that could be achieved.

- Regulators and policy makers were hopeful for future possibilities of sharing data with industry (whilst respecting commercial in confidence information) and having industry share information with them.

Appendices

Appendix 1. Questionnaire used in survey

Appendix 2. Organisations who participated in survey

References

1. Jackson, B.R., et al., *Implementation of Nationwide Real-time Whole-genome Sequencing to Enhance Listeriosis Outbreak Detection and Investigation.* Clin Infect Dis, 2016. **63**(3): p. 380-6.

2. Ribot, E.M. and K.B. Hise, *Future challenges for tracking foodborne diseases: PulseNet, a 20-year-old US surveillance system for foodborne diseases, is expanding both globally and technologically.* EMBO Rep, 2016. **17**(11): p. 1499-1505.

3. CDC, *Whole genome sequencing improves the detection and investigation of foodborne outbreaks*. https://www.cdc.gov/listeria/surveillance/whole-genome-sequencing.html 2017.

Appendix 1: Questions

MLA NGS Strategy Project (MLA PROJECT NO. V.MFS.0415)

Parameters for the information gathering stage of the project - these will be asked of a limited number of regulators and/or testing labs, both national and international. Information will also be gathered by searching web sites and public information.

Obviously, there may be some areas deemed too sensitive or participants are unwilling to discuss, so this can just be noted.

Focus:

- Generally interested in NGS related matters as they pertain to food isolates and raw products (not really focused on ready to eat foods perhaps with the exception of small goods containing red meat). I suspect that the specifics of the food type will not be a big factor in the discussion (just something to keep in mind).
- Expect primary focus to be on STEC and *Salmonella*

Areas of interest

- Metadata collection (e.g., food source, date, location). Possible interest in whether all or some of this data will remain attached to any DNA sequence data that will be created. There would be some interest on the extent to which this sample specific data would be disclosed to various parties (e.g., the general public, shared with regulatory jurisdictions, industry for research).
- Data ownership and storage. This applies to both genomic DNA sequence data / SNP data and sample metadata
    - Issues around the large quantity of data and how to safely store.
    - Are there issues with cloud / off-shore storage?
    - Who will have access?
    - How long will data reside in the database (12 month rolling dataset, indefinite, etc.)
- Standards and accreditation that might be applied to the use of NGS technologies for typing. Interest in quality assurance around both the DNA sequence data itself and the data analysis tools
- Likely schedule for the switch from PFGE to NGS - SNP analysis.
- If known, what SNP methods or data analysis pipelines are/will be used.
- Has any industry engagement around NGS testing planned by labs/regulators for the near future
- Future directions: what are labs moving into i.e. single molecule sequencing, long reads or metagenomics rather than isolates
- Turnaround time for performing NGS on food isolates i.e. is it currently real time or retrospective

Overall question to be answered by this data (not to be asked)

- How uniform is the NGS technology and analysis tools being implemented here in Australia compared to the rest of the world

Appendix 2: Organisations Contacted

Organisation	Type	
Agrifood Technology	Commercial Detection Laboratory	
ALS Food & Pharmaceutical	Commercial Detection Laboratory	
Analytical Microlabs	Commercial Detection Laboratory	
Australian Food Corporation	Commercial Detection Laboratory	
Biotech Laboratories	Commercial Detection Laboratory	
Biotest Laboratories	Commercial Detection Laboratory	
DTS Food Laboratories	Commercial Detection Laboratory	
Fletcher International	Commercial Detection Laboratory	
Greenham & Sons	Commercial Detection Laboratory	
JBS	Commercial Detection Laboratory	
John Dee	Commercial Detection Laboratory	
Merieux Nutrascience Pty Ltd	Commercial Detection Laboratory	
Midfield Meat International	Commercial Detection Laboratory	
Nav Labs Pty Ltd	Commercial Detection Laboratory	
Northern Co-Operative Meat	Commercial Detection Laboratory	
Oakey Beef Exports	Commercial Detection Laboratory	
ProMicro Pty Ltd	Commercial Detection Laboratory	

Rivalea (Australia) Pty Ltd	Commercial Detection Laboratory	
SA Analytical Laboratory Services	Commercial Detection Laboratory	
Stanbroke Beef Pty Ltd	Commercial Detection Laboratory	
Symbio Laboratories	Commercial Detection Laboratory	
Thomas Food International	Commercial Detection Laboratory	
Wondonga Rendering	Commercial Detection Laboratory	
Forensic & Scientific Services Qld	Public Health Reference	
ICPMR, NSW	Public Health Reference	
Microbiological Diagnostic Unit Vic	Public Health Reference	
PathWest WA	Public Health Reference	
South Australia	Public Health Reference	Did not respond
Tasmania	Public Health Reference	
Department of Agriculture and Water Resources	Regulator	
Health Protection Service ACT	Other	
OzFoodNet ACT	Regulator	
OZFoodNet Commonwealth	Regulator	
OzFoodNet Qld	Regulator	
Safe Food Qld	Regulator	
Northern Territory	Regulator	Did not respond
National Centre for Epidemiology & Population Health, Australian National University	Other	