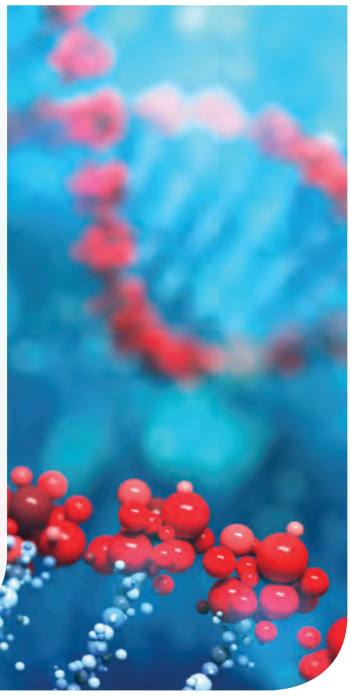
Chief Scientific Adviser's
Science Report

Issue Three: Whole-genome sequencing of Foodborne Pathogens

Food Standards Agency
food.gov.uk

"My last two reports focused on microbiological and chemical risks, and the expert advice we receive from our scientific advisory committees. This report focuses on the exciting developments in whole-genome sequencing, and how this powerful and rapidly developing technology is being increasingly utilised by the FSA."

Professor Guy Poppy,
FSA Chief Scientific Adviser



The ability to sequence whole genomes allows scientists to identify, and differentiate between, different bacterial and viral strains. The increasing speed and decreasing costs of this rapidly evolving technology has had a transformative effect on our ability to investigate foodborne disease outbreaks. The improving level of detail we're able to operate in, will allow us to quickly identify and control the source of outbreaks, which is an increasingly difficult challenge in the globalised food system.

As well as discussing the benefits that developments in whole-genome sequencing brings to a range of competencies across the FSA's remit, this report identifies some of the key areas where the FSA, in collaboration with partners, is supporting future advancements in the technology – we are helping drive the technology as well as using it to improve public health.

We also report on our growing research interests in the human microbiome, another area that has recently seen critical and far-reaching scientific development. Understanding how changes to the microbiota in our gut can shape immune responses to both physiological and pathological conditions will be central to how we engage with health practitioners and the food industry and has the potential to revolutionise food and public health in the near future.

This report showcases the huge benefit that scientific and technological innovation can bring to the FSA's work. Such innovation is critical in allowing us to fulfil our duty to ensure that people can trust that the food that they eat is safe and honest.

Lead Analysis: Whole-genome sequencing

When completion of the first draft of the human genome (the entire genetic make-up of a human – in other words our complete DNA sequence) was announced by Tony Blair and Bill Clinton in 2000, it represented more than ten years of work from a huge international consortium, at a cost of billions of pounds. DNA sequencing technology has moved on dramatically since then, with the development of entirely new technologies for sequencing DNA (collectively termed 'next generation sequencing'). These new technologies have revolutionised the field, enabling whole genomes to be sequenced very rapidly and at a fraction of the cost (see 'Next Generation Sequencing').

The power of whole-genome sequencing (WGS) to reveal the entire genetic makeup of a microorganism allows us to gain an unprecedented understanding of the similarities and differences between species, but also between different individuals of the same species. This means that we can look at the differences between individual bacterial and viral strains with a precision that was not previously possible or practical; and this ability is transforming, among other things, the investigation of foodborne disease outbreaks.

New technologies have revolutionised the field, enabling whole genomes to be sequenced very rapidly

Whole-genome sequencing

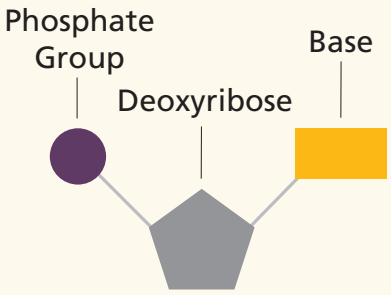
The structure of DNA

Genetic information – the information that can be passed down from one generation to another – is (in most organisms) carried in the form of a molecule called deoxyribonucleic acid (DNA). DNA is composed of units called nucleotides, strung together in a row. There are four different nucleotides (each with a different 'base') found in DNA, represented by the letters A, T, C and G; the sequence in which those nucleotides appear in the DNA holds an individual's genetic information. Some viruses use a slightly different molecule (ribonucleic acid, or RNA) to encode their genomes, which must be converted to DNA before it can be analysed.

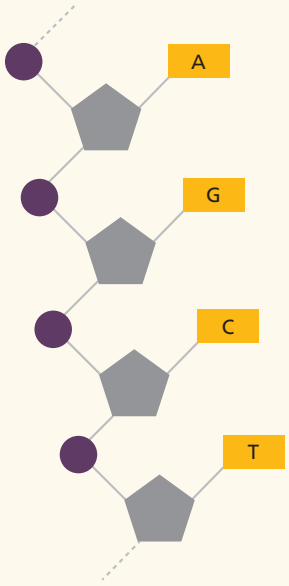
Whole-genome sequencing

Whole-genome sequencing involves 'reading' the order of all the nucleotides, one by one; in the case of the human genome, there are about three billion nucleotides to read. The genomes of bacteria and viruses are much smaller than that – typically a few million nucleotides for bacteria and just a few thousand for viruses – and can now be sequenced for less than £100, and in just a few hours.

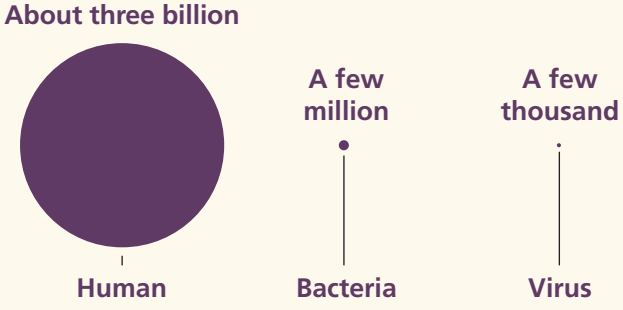
Nucleotide structure



Nucleotides combine to form DNA



Comparing the number of nucleotides in genomes



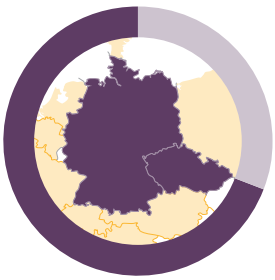
An international *Salmonella* outbreak

In June 2014, Public Health England (PHE) was alerted to an outbreak of salmonellosis in a hospital in central England. The investigations that followed provide an excellent illustration of how PHE and the Food Standards Agency (FSA) can make use of WGS to rapidly find the source of an outbreak and prevent further cases of illness.

Salmonella enterica is a bacterium that is present in many domestic and wild animals, and can cause food poisoning in humans.

Salmonella cases

cases between May and September 2014



69% of cases linked to eggs from a producer based in Germany and Czech Republic

Human infection is usually through the consumption of contaminated food, and in a small proportion of cases can have serious consequences including hospitalisation and death.

FSA research indicates that *Salmonella* alone accounts for more than 10,000 GP consultations and more than 2,500 hospital admissions each year in the UK.

The strain of *Salmonella enterica* causing the outbreak was initially analysed using two routine traditional techniques, serotyping and phage typing, and determined to be *Salmonella* Enteritidis PT14b. Further analysis used a technique called multi locus variable number tandem repeat analysis (MLVA) to look at nine particular regions of the genome, and cases that were part of the outbreak were defined as those that had exactly the same MLVA profile, or that differed in just one of the regions.

See 'Traditional techniques for distinguishing bacterial strains' for more information on the techniques.

Further outbreaks of *S.* Enteritidis PT14b with a matching MLVA profile were detected in the

Northwest and South of England; at the same time, outbreaks were also reported in France and Austria, all with the same MLVA profile.

Investigations were carried out collaboratively between PHE, the FSA and other public health bodies including local authorities, to identify the source of infection and implement control measures in order to prevent further cases. In total, 287 people became unwell from May to September 2014. Eggs from a German egg producer had been implicated in the French and Austrian outbreaks, and food chain investigations showed that 198 of the cases in England and Wales could be linked to eggs from the German egg producer. The MLVA profile of S. Enteriditis PT14b isolated from eggs from this company matched that from the UK cases.

WGS provided important corroborating evidence linking the outbreaks of infection to eggs from the company. *Salmonella* isolated from clinical and environmental samples, including eight isolates from eggs from the company, were sequenced alongside the other investigations. Analysis of all 332 isolates showed that they clustered together – in other words, all of the *S*. Enteritidis PT14b isolates that were associated with the outbreaks were much

Traditional techniques for distinguishing bacterial strains

At present, several different techniques are used to identify and characterise bacteria, and to distinguish between different strains of a species. The method of choice varies depending on the species in question and on precisely what information is required. Some of the commonly used methods are:

Serotyping – different bacterial strains have different sets of molecules coating them; antibodies are used to determine which molecules are present on the outer surface.

Phage typing – distinguishing between strains based on their susceptibility to infection by various types of bacteriophage (a group of viruses).

Pulsed Field Gel Electrophoresis (PFGE) – the bacterial genome is chopped up at specific points, to create fragments that are then separated on a gel according to their size. Different strains will have different patterns of fragments.

Multi Locus Variable Number Tandem Repeat Analysis (MLVA) – particular regions of DNA are very repetitive, and the number of repeats of a sequence in a particular region varies between different bacterial strains. A few such areas are analysed to determine how many repeats there are in each.

While these various techniques can furnish public health officials and clinicians with very helpful information, particularly when used in combination, they do not have the exquisite resolving power that WGS has. Furthermore, maintaining the resources (including staff expertise) to perform these and other assays is a challenge for cost efficiency and resilience in public health laboratories. WGS in principle offers one simplified approach, easily reproducible across different laboratories.

more closely related to each other than they were to *S*. Enteritidis PT14b from sources not linked to the outbreak.

Although the other investigations had pointed in the right direction, the exquisite level of detail about the various *Salmonella* isolates from WGS provided a clear genetic link between the outbreak cases, the premises that had been implicated and eggs from the German producer. This enabled a far more detailed understanding of how the *Salmonella* had spread to different countries in Europe and provided compelling evidence that the UK cases were linked to the German eggs.

Next Generation Sequencing

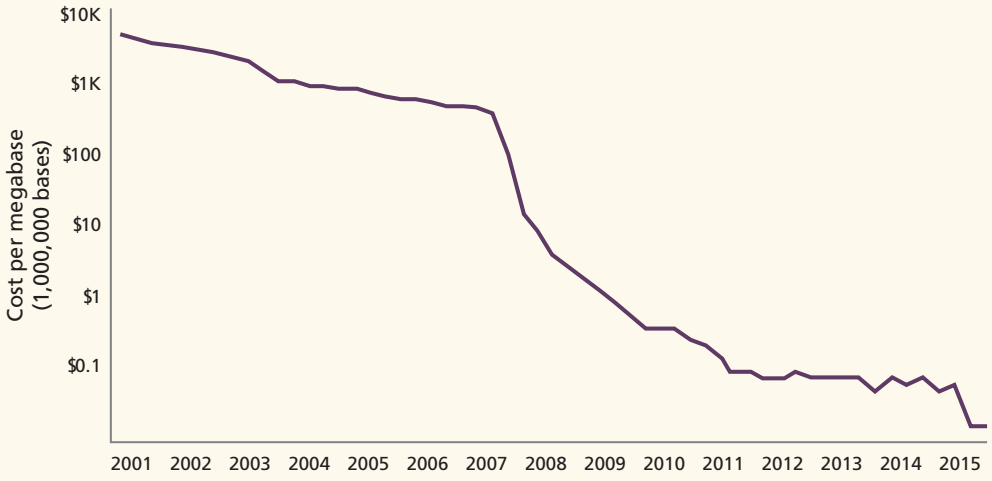
For many years, DNA sequencing was done more or less the same way, based on the technique developed by Nobel Prize-winning British biochemist Fred Sanger in the 1970s.

But recently, the development of several new technologies – broadly referred to as 'next generation sequencing' – has led to huge increases in the speed at which sequencing can be performed, along with decreases in the cost.

Several companies have developed different approaches, but in general they all have a major benefit: unlike Sanger sequencing, many millions of DNA molecules can be sequenced at the same time. This is largely what has precipitated the greatly increased speed and, in turn, the decreasing cost.

As these technologies are developed further, and additional new technologies come onto the market, even more improvements to speed and cost are likely. But an added benefit is that the size of the instruments used to perform the sequencing has reduced markedly. At least one company has even developed a product so small that it can fit into your pocket, enabling samples to be sequenced at source, getting rapid results.

Figure 1 – Cost per raw megabase of DNA sequence



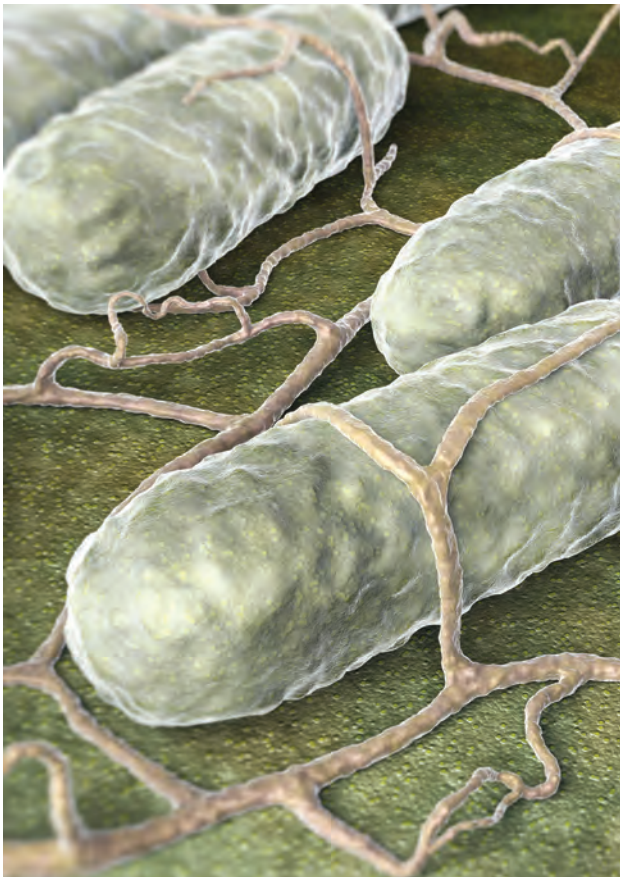
Source: National Human Genome Research Institute

Benefits of WGS

WGS has a multitude of uses, and the potential to change the way we work and improve public health in many ways. Some of these are outlined below.

Resolution and speed in outbreak investigations

The use of WGS in the *Salmonella* outbreak described above demonstrates its power in confirming the source of an international outbreak. But this is by no means the limit – it is easy to see how, in the future, WGS could become the primary tool for obtaining data to identify, characterise and compare strains in outbreak investigations. PHE have made very important contributions in this developing area, and after successful trials will have replaced the traditional microbiological methods for distinguishing strains with WGS, for routine analysis of a variety of foodborne bacteria: *Salmonella*, Shiga toxin-producing *E. coli*, *Campylobacter* and *Listeria*.



Salmonella

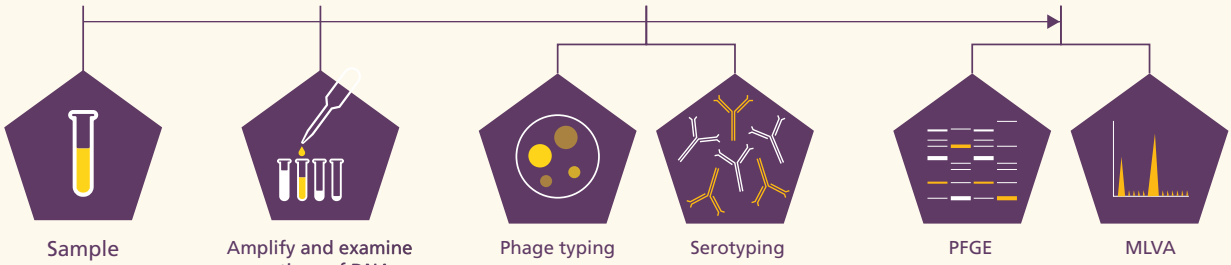
Not only does the level of detail provided by WGS greatly enhance our ability to tell the difference between similar strains, it also provides information about the evolutionary relationships between the strains (ie how closely related they are to each other, and how they are changing over time). This gives a much better ability to detect clusters and identify outbreaks that might otherwise have been thought of as independent, unlinked sporadic cases – and this in turn helps to trace the source and prevent further cases. For example, a case of listeriosis in 2012 prompted an investigation that used WGS as evidence to identify a further thirteen cases associated with consumption of the same product, going back over the previous two years. As a result, control measures were taken to prevent further illnesses.

WGS could become the primary tool for obtaining data to identify, characterise and compare strains in outbreak investigations

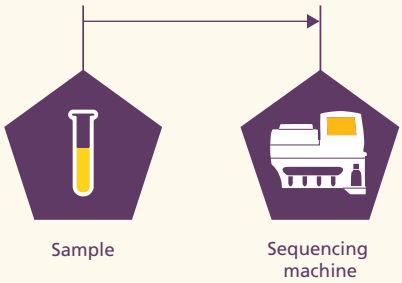
As well as the benefits that come with such detailed information, WGS can greatly speed up investigations. For example, traditional methods for investigating *Salmonella* involve a series of steps, some of which are very time-consuming (see 'Traditional techniques for distinguishing bacterial strains' for information on some of these steps).

Speed of WGS vs traditional steps

Traditional method



WGS method



By replacing these steps with WGS (which provides the necessary information for identification and typing, but gives additional detail), the turnaround time for the analysis of *Salmonella* isolates has been reduced from around 20 days to around a week. And a much faster investigation process means that the source of outbreaks can potentially be found more quickly and measures put in place to stop further cases – possibly saving lives.

Analysis of new emerging pathogens

Should a new variant of a pathogen – or indeed a completely new pathogen – emerge, WGS could be crucial for early detection. WGS could also be instrumental in analysing and understanding it. By comparing the sequence of a new emerging pathogen to known sequences, it will be possible to rapidly gain a basic understanding of the biology of the organism – for example by looking for genes encoding known virulence factors (such as molecules that help the organism invade or colonise the host, or toxins).

Use as a monitoring tool

WGS can be a powerful tool for monitoring and surveillance. In the USA, for example, WGS has been used for *Listeria monocytogenes* since 2013, and this helped

Correct interpretation of the data requires the skilled expertise of bioinformaticians

to identify cases of listeriosis that were part of a multi-state outbreak linked to three different ice cream production facilities owned by the same company. Because of its precision, WGS provides very useful forensic evidence for source tracking and regulatory action.

Authenticity

Authenticity – ensuring food is what is says it is – is another area where WGS has potential uses. In 2013, the detection of horsemeat in products purporting to be beef highlighted the need for 'speciation' testing. Traditional DNA-based methods (eg real-time PCR) can be very powerful in detecting the presence of particular species in a given food sample, in order to ensure that foods have not been adulterated. However, the limitation is that this is a targeted approach: you have to look for something in particular. Next generation sequencing technologies, however, have the potential to tell us all the species that are present in a particular sample – a so-called unbiased approach – enabling detection of even completely unexpected species in a sample. The limitation, though, is that these technologies tend to be less sensitive than traditional DNA-based methods.

Challenges

In order for WGS to fulfil its potential, sharing of data is of paramount importance: knowing the sequence of a particular bacterium or virus is of no use without a database of related genome sequences to compare it with. However, this raises questions about ownership of samples and the data from them, and about intellectual property rights – which may lead to some resistance to sharing data freely. An additional challenge in this area is that for data from different sources to be compared meaningfully, there will need to be globally standardised, quality-controlled procedures and ways of recording data. And crucially, patients must be safeguarded from any possible breaches of confidentiality of privacy resulting from samples being traced to them. This illustrates the increasingly important role that data will continue to play, both in the utility and responsible management of large datasets. The emergence of Big Data, and how it will play a crucial role in the FSA's work, will be covered in a later CSA report.

Another big challenge in implementing the use of WGS more widely is how to handle the huge amount of data generated – how to store it, and how to interpret it. Correct interpretation of the data requires the skilled expertise of bioinformaticians, and thus it will be crucial to ensure that there are enough people trained in this discipline.

These challenges are by no means insurmountable, but will need to be carefully considered and navigated by international collaborative groups – such as the GMI project that the FSA has long-standing involvement with (see 'International Developments').

Another big challenge in implementing the use of WGS more widely is the huge amount of data generated – how to store it, and how to interpret it

International developments

Countries involved in GMI

International cooperation and sharing of data is key to unlocking the full potential of WGS. Realising this, the FSA became involved at an early stage with a major international collaborative effort in this area. The Global Microbial Identifier (GMI), as its name suggests, is an initiative on a truly global scale, and the FSA has been a member of its steering committee since 2012. It originated as a group of 25 participants, but has grown to involve over 200 experts from a broad range of disciplines (including microbiology, bioinformatics, epidemiology and public health), and representing over 40 countries.

The GMI is working towards having one single global system of databases for sharing the genomic sequence data from disease-causing microorganisms. This would provide an incredibly powerful resource, linking each genome sequence to metadata such as epidemiological details.

By placing ourselves at the heart of this and other international efforts, the FSA is helping to make sure that the resources being developed will fit the UK's needs,

The FSA is helping to make sure that the resources being developed will fit the UK's needs, and that the UK is in the best possible position to make use of them

isolates have been sequenced to date

sequenced per month

and that the UK is in the best possible position to make use of them. We have also joined forces with Fera to fund a research fellow, with the express aim of scoping and developing capability in genomic sequencing for epidemiology of microbial contaminants in the food supply chain.

In addition to interactions with UK-based stakeholders and the GMI, the research fellow has forged strong links with the United States Food and Drug Administration, particularly in relation to their GenomeTrakr network. This is a network of laboratories (including three in the UK) that collect genomic data from foodborne pathogens and make the sequences and associated geographical data publicly available. This is no small undertaking, with more than 51,000 isolates sequenced to date, and over 1,000 regularly being sequenced each month. Learning from the experiences of the GenomeTrakr system will help us with our future endeavours.

Developing the potential through funding research

In 2012, the FSA held an expert workshop to explore the opportunities and challenges that new molecular biology technologies brought in relation to foodborne disease outbreaks. Following enthusiasm at this workshop for the potential of next generation sequencing technologies to bring public health benefits, and the recommendation that "*High-throughput sequencing is currently capable of providing a significant benefit to outbreak investigations and should be used from the next outbreak onwards*", the FSA has funded several projects to develop the potential of WGS; some examples are given below.

One important pilot study in 2013 involved sequencing isolates of the disease-causing bacterium *Listeria monocytogenes* from red meat, as well as several other closely-related *Listeria* species. Analysis of these sequences using comparative bioinformatics to understand the relationships between individuals of the same species, and between species, demonstrated that WGS could successfully be used for tracing and source attribution. This project fed into a wider international 'proof of principle' study, led by the Global Microbial Identifier initiative.

Another project brought together an interdisciplinary consortium comprised of the Institute of Food Research (now named the Quadram Institute), The Genome Analysis Centre, the University of East Anglia, PHE, the University of Swansea and the University of Liverpool, to focus on another bacterium: *Campylobacter coli*. Although this species is thought to be responsible for around 15% of human *Campylobacter* infections (equating to about 75,000 cases per year), there was a dearth

of data on its genome sequence. This project filled that gap, generating sequence data (and associated data) on *C. coli* isolates from a range of sources, and paving the way for studies on environmental reservoirs and transmission routes, and improved diagnostics, tracking and source attribution of *C. coli* in an outbreak.

A project jointly funded by the FSA and the Biotechnology and Biological Sciences Research Council involved sequencing *Campylobacter* strains isolated at key stages from poultry processing through to human disease. Poultry is the bacterium's main reservoir; the data from this project were used to gain insights into the genetic basis for some strains being able to survive poultry processing and infect humans. Ultimately, this approach may help to identify targets for interventions to eliminate or reduce *Campylobacter* contamination of poultry.

Most recently, we have invested in research to explore innovative new ways to utilise WGS data. Two projects that are currently running are looking at possible ways to integrate sequence data with other associated data, such as geographical origin, to improve food safety outcomes.

Conclusions

WGS is an incredibly powerful technique: it is fast, precise and (if implemented widely) cheap. It is already being utilised by PHE and the FSA to aid outbreak investigations; but it has a multitude of other potential applications in relation to food. Maintaining our central role in international efforts to arrive at European and global frameworks to support the best possible use of WGS is crucial for fostering an environment in which WGS can be used to our best advantage.

While traditional microbiological, epidemiological and food chain investigation methods still have an important role to play, there is no doubt that WGS is revolutionising outbreak investigations, and is likely to have similar impact in other areas of interest to the FSA. It will be a vital tool for moving closer to a global 'one health' approach to preventing and mitigating the effects of diseases that emerge at the interface between humans, animals and their environments. It is important that the FSA use cutting-edge technology and collaborate with the best partners across the globe to obtain the best evidence available for ensuring UK consumers can eat safe and honest food.

Also of interest...

The microbiome

The human microbiome is formed by about 100 trillion microbes including eukaryotes, archaea, bacteria and viruses and accounts for 1 to 3 percent of our body mass. These microbes are generally non-pathogenic and co-exist symbiotically with us, providing in healthy individuals a number of metabolic functions that humans lack.

Since the late 1990s an increasing number of research projects, including the Human Microbiome Project, a $115 million programme launched in 2008 by the National Institutes of Health in the U.S., have been investigating how changes in the human microbiome are associated with human health or disease.

It is now known that the microbiota, specifically those found in our gut, are able to shape intestinal immune responses both in physiological and pathological conditions. Experts agree that the dramatically rising rates of allergic, autoimmune and other inflammatory diseases are associated with the fact that our indoor-centred way of living and a diet rich in processed foods have depleted our microbiota. Specifically, scientists have identified the loss of biodiversity in the microbiota as central in the high rates of inflammatory diseases in the developed world. When it comes to bacteria biodiversity, a healthy diet is the central ingredient.

The Quadram Institute in Norwich and its £80 million investment illustrate the level of interest in this exciting area. The FSA, in collaboration with this institute, held a workshop for the Advisory Committee on the Microbiological Safety of Food to consider the microbiome and to explore innovative research into the role the microbiome might play in resilience/vulnerability to bacterial and viral foodborne disease. This is an opportunity to be at the forefront of an emerging discipline of health and food research. The work will look at developing and implementing next generation sequencing based tools for a step change in the speed and accuracy of diagnostic testing, allowing the investigation of all organisms present in samples. Understanding the data, through expanding our ability to interrogate microbial communities, and presenting useful information to health practitioners and the food industry seems a rational approach which would allow investigations to go right through the food chain.

The FSA is also developing a programme of research on adult food allergy to try to understand the causes behind the rise of adult allergies over the past 15 years. The FSA is including microbiome experts and commissioning lines of research in the microbiome in this programme of work. Our progress in this exciting and innovative area will be covered in a future CSA report.

Acknowledgements

With thanks for the contributions of Jesús Alvarez-Piñera, Mariam Orme and David Self.

Further Reading

For further information on whole-genome sequencing of foodborne pathogens, please see:

COMPARE http://www.compare-europe.eu/

Dallman, Byrne, Ashton, Cowley, Perry, Adak, Petrovska, Ellis, Elson, Underwood, Green, Hanage, Jenkins, Grant and Wain (2015). Whole-Genome Sequencing for National Surveillance of Shiga Toxin-Producing Escherichia coli O157. Clin Infect Dis, 61 (3): 305-12.

Food Standards Agency. The Application of Molecular Epidemiology to Investigations of Foodborne Disease Outbreaks: Current Status and Future Plans. Report of a Workshop held 17th January 2012, London, UK http://www.food.gov.uk/sites/default/files/multimedia/pdfs/ngsworkshop.pdf

FS101072. FSA project report. Application of whole genome sequencing to fully characterise campylobacter isolates from IID1 and IID2 studies. http://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/fs101072

FS101087. FSA project report. Generating tools for the molecular epidemiology of Campylobacter coli by next generation genome sequencing. http://www.food.gov.uk/science/research/supportingresearch/strategicevidenceprogramme/x02projlist/fs101087

FS102029. FSA project report. Whole genome molecular epidemiology of E.coli O157 isolates from humans, food and the environment. http://www.food.gov.uk/science/research/foodborneillness/fs102029

FS102062. FSA project report. Listeria Genomics Pilot Project. http://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/fs102062#toc-5

FS231043. FSA project report. Extension to the IID2 study: identifying the proportion of foodborne disease in the UK. https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/fs231043ext

FS231086. FSA project report. Modelling campylobacter survival and spread through poultry

processing: a population genomics approach. http://www.food.gov.uk/science/research/foodborneillness/b15programme/b15projects/fs231086#toc-6

Global Microbial Identifier http://www.globalmicrobialidentifier.org/

Inns, Lane, Peters, Dallman, Chatt, McFarland, Crook, Bishop, J Edge, Hawker, Elson, Neal, Adak and Cleary (on behalf of the Outbreak Control Team) (2015). A multi-country *Salmonella* Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014. Euro Surveill., 20 (16): pii=21098.

Luheshi L, Raza S, Moorthie S et al. PHG Foundation (2015). Pathogen Genomics Into Practice. ISBN 978-1-907198-18-2. http://www.phgfoundation.org/reports/16857/

U.S. Food and Drug Administration Whole Genome Sequencing (WGS) Program http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/default.htm

For further information on the microbiome, please see:

NIH Human Microbiome Project http://hmpdacc.org/

Round, J. L.; Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. Nature Reviews: Immunology 9 (5): 313–323.

For more information, visit: **food.gov.uk**

Let's keep connected at **food.gov.uk/facebook**

Join our conversation **@foodgov** using **#CSAreport**

Watch us on **food.gov.uk/youtube**

Repin us at **food.gov.uk/pinterest**

Revine us at **food.gov.uk/vine**

Published by the Food Standards Agency 2016