Campden BRI
Station Road
Chipping Campden
Gloucestershire
GL55 6LD, UK

Tel: +44 (0)1386 842000
Fax: +44 (0)1386 842100
www.campdenbri.co.uk

Microbial Whole Genome Sequencing and the Food Industry

Whole Genome Sequencing (WGS) is a technique that enables the complete DNA base sequence of a microorganism to be determined. As the DNA sequence is virtually specific to an individual cell within a population, it is a technique that can be used to identify an individual very precisely and differentiate it from other very similar, but different organisms.

This white paper discusses the potential applications of whole genome sequencing to the food industry.

Contact:

Greg Jones
Greg.jones@campdenbri.co.uk
+44(0)1386 842143

Issued: February 2017

Introduction

WGS has the potential to render other forms of microbiological identification obsolete. It is more accurate than a serotype, more discriminatory than a pulsed-field gel electrophoresis assay and it can prove relationships between strains with higher resolution than ever before. This is the method which has been adopted by regulatory agencies such as Public Health England and the Food and Drug Administration (FDA) in the USA to identify food poisoning outbreaks and link isolates from outbreaks with those from foods or environmental sources.. Food companies are starting to become more aware of this area and wish to have a constructive dialogue with government agencies; however, there can be a lack of knowledge regarding the technology and the potential uses of it in an industrial setting. Table 1 gives examples of how WGS has been used in the last few years.

Table 1: Examples of WGS use since 2014

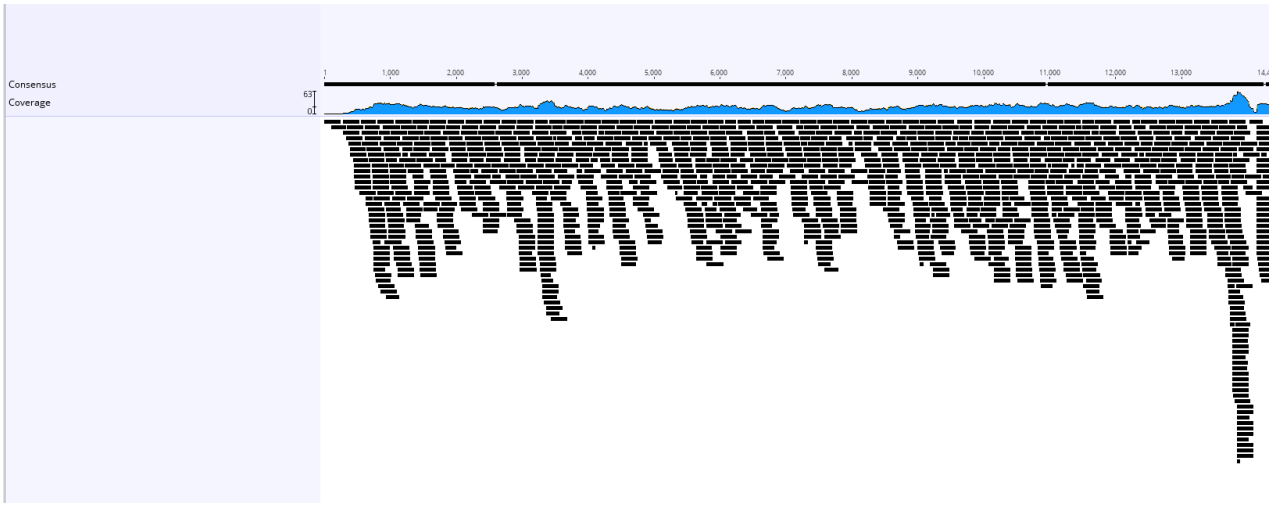
Use	**Summary of work**	**Reference**
Strain differentiation	Epidemiology and MLST used to differentiate apparently clonal *Listeria monocytogenes* isolates associated with cheese and delicatessen meat.	(Schmid et al., 2014)
Strain tracking	*L. monocytogenes* strains tracked across states and localised to source.	(Stasiewicz et al., 2015)
Outbreak investigation	Outbreaks of *Escherichia coli* poisoning were differentiated more exactly, and smaller outbreaks detected.	(Dallman et al., 2015)
Brewing spoilage organism investigation	Genetic differences between spoilage and beneficial strains of *Brettanomyces* (*Dekkera*) in the brewing industry discovered.	(Crauwels et al., 2014)
Crop genetics	Wild and domesticated Capsicum pepper evolution tracked and genetic differences explored with a view to improve future varieties.	(Qin et al., 2014)
Serotyping	*Salmonella* serotyping method was developed using curated databases of serotype-specific genes.	(Zhang et al., 2015)
Antimicrobial resistance	Investigation into plasmid-based AMR genes in *E. coli*.	(Losada et al., 2016)

Method

The technique has become more prevalent in the last five years due to rapid advances in sequencing technology that have led to dramatic falls in cost of sequencing. It is now possible to sequence microbial genomes on a routine basis for a few hundred pounds each. Regardless of the technique used, the generation of huge amounts of sequence data has become entirely unremarkable. When a genome is sequenced, the initial output is hundreds of thousands of short sequences. Each of these sequences is a few hundred bases long and represents a tiny fragment of the total genome. The next challenge is to assemble these reads by comparing them to each other and ordering them according to

their overlapping ends. This process is analogous to reconstructing an ancient document from fragments of parchment.

Figure 1: A small portion of a genome assembled from raw sequence reads. Each of the small black bars represents one sequence approximately 500 bases long



The workflow to assemble a genome is straightforward:

1. Obtain an isolate via culture-based methods.
2. Extract the DNA.
3. Prepare the DNA for sequencing ("Library Preparation").
4. Run the sequencer.
5. Assemble the short raw reads into longer sequences using software.

Pathogen identification

For pathogen identification, the next step is to give some meaning to the sequence by comparing it to other sequences to assess their similarity. The comparison is reliant on the number of other genomes against which your submitted sequence is compared. The analysis gets larger as more genomes are added in to the comparison. Comparing whole genomes against databases of other whole genomes is currently performed by the FDA in the USA using their 'Genome Trakr' service (Allard et al., 2016). This service relies on the vast storage and computing power available to them from the National Centre for Biotechnology Information (NCBI). This is a resource available to anybody, and a genome submitted for analysis will be placed into context via comparison against other sequences. The output is a phylogenetic tree similar to the one shown in Figure 2.

Figure 2: Output from Genome Trakr (https://www.ncbi.nlm.nih.gov/pathogens/)

In the example above, the genome labelled as an 'Environment/Food' sample is highlighted in red, and is shown to cluster very closely with a set of isolates designated as 'Clinical'.

The necessary use of a public database for this analysis has led to concern from some in the food industry who fear that doing the right thing and submitting sequences will reflect badly on them in the event that their sequence is shown to be related to sequences obtained from food poisoning outbreaks. Despite assurances from the FDA in a recent meeting, US industry representatives still have some concerns that sequences submitted with accompanying descriptions of source, that could ultimately be traced back to the company of origin, are so sensitive, that some companies would not wish to submit at all. The mood in the UK industry is similar, with companies approaching this method with a degree of caution.

Is this caution warranted? A similar tool for tracing outbreaks exists in the form of PulseNet (http://www.cdc.gov/pulsenet), based on DNA fingerprinting technology. What is new for WGS is the finer level of discrimination. As this information is available to anyone, the submitting company will be alerted to any clinical link at the same time as the regulator, allowing earlier action to be taken. Submitters' names are not made publically available, but could be held by the regulator. Industry is therefore more likely to submit sequences if their describing metadata can be made anonymous. Earlier notification of a link to an outbreak is in everyone's best interest, and it will be in the submitter's interest to be removed from the investigative focus if the submitted sequence does *not* match clinical data. Despite these clear advantages, there is still the worry that a current isolate can be linked to outbreaks that occurred at any time in the past or present, and that a current outbreak could be linked to an isolate submitted at any time. If the food industry is to use this technique and work constructively with the regulators, these issues need to be addressed and binding assurances given by the regulator that the industry's desire to protect public health through the use of WGS will not result in an increased probability of prosecution should an unfounded link be made.

A recent paper from the FAO stated:
"Depending on the country situation, adoption of WGS by the food industry may either precede or follow its adoption by regulatory agencies, and may incur additional challenges and concerns including legal/liability issues. It is important that governments consider providing an enabling environment for the food industry to generate, use and share relevant WGS data. In this way, public-private collaboration would promote applications of WGS in the most desirable and appropriate way for improved national food safety management." http://www.fao.org/3/a-i5619e.pdf

Campden BRI is actively working with industry and regulators to advise and help reach a mutually beneficial result. If you would like to explore these options in more depth, please get in touch.

Non-public uses of WGS

Genome sequencing can be used as a more contained solution to managing pathogenic organisms. Techniques are available that allow a controlled, in-house comparison of genome sequences to be made. For example, a manufacturer will be able to compare environmental isolates of *Listeria* against isolates found in food. This confidential, local analysis and the retention of sequences will allow the relationships between strains to be monitored over time and answer questions regarding the source of isolates as they are added to the dataset. In the event of an isolate being found in a final product, the question of whether it originated in raw material or the manufacturing environment could be answered. Data such as this can also be compared to Riboprinter data via *in silico* fragmentation of the 16S ribosome gene from a whole genome sequence.

Another major step forward with WGS is the potential to identify and track specific genetic elements, such as antibiotic resistance genes. It is now possible to predict the compounds a particular isolate is resistant to from WGS data (Day et al., 2017). The implications here are particularly relevant to meat production in light of the current concern surrounding use of antibiotics. This may lead to new microbiological specifications in the future, not only specifying the levels of organisms that are acceptable, but also the range of genetic elements the microflora must not harbour.

At the time of writing, the major bottleneck in any analysis is the amount of computing power and analysis software available to the user. The software currently available requires specialist knowledge to be able to operate it properly. This can be frustrating for the general user who does not need to acquire such knowledge as part of their usual role. The solution to this is for the sequencing and analysis to be outsourced to a body such as Campden BRI. The food industry thus has a unique resource available that combines expertise in the technicalities of sequencing and analysis with specific food-industry knowledge. Campden BRI currently offers WGS as a stand-alone tool to answer specific questions, and as part of our Microbial Identification service. For example, a *Salmonella* isolate can now be confirmed as such using culture and MALDI-TOF with WGS used to identify the serotype.

Conclusion

WGS is a tool that will give the food industry greater understanding of the individual organisms that make up the food itself and its accompanying microflora. The ability to trace pathogenic organisms to source is being utilised by Public Health bodies, and it should be used by the Food industry once concerns over the content of associated metadata have been allayed. WGS also has the potential to reveal the effects of up-stream strategies such as antibiotic use on the microflora of a product. Campden BRI are well positioned to lead the food industry into the area of WGS. The service offered combines the latest developments in the field with the deep understanding of the food industry the company has to offer.

Reference s

Allard,M.W., Strain,E., Melka,D., Bunning,K., Musser,S.M., Brown,E.W., and Timme,R. (2016). The practical value of food pathogen traceability through building a whole-genome sequencing network and database. Journal of Clinical Microbiology JCM-00081.

Crauwels,S., Zhu,B., Steensels,J., Busschaert,P., De Samblanx,G., Marchal,K., Willems,K.A., Verstrepen,K.J., and Lievens,B. (2014). Assessing genetic diversity among *Brettanomyces* yeasts by DNA fingerprinting and whole-genome sequencing. Applied and Environmental Microbiology *80*, 4398-4413.

Dallman,T.J., Byrne,L., Ashton,P.M., Cowley,L.A., Perry,N.T., Adak,G., Petrovska,L., Ellis,R.J., Elson,R., and Underwood,A. (2015). Whole genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. Clinical Infectious Diseases civ318.

Day,M., Doumith,M., Jenkins,C., Dallman,T.J., Hopkins,K.L., Elson,R., Godbole,G., and Woodford,N. (2017). Antimicrobial resistance in Shiga toxin-producing *Escherichia coli* serogroups O157 and O26 isolated from human cases of diarrhoeal disease in England, 2015. Journal of Antimicrobial Chemotherapy *72*, 145-152.

Losada,L., DebRoy,C., Radune,D., Kim,M., Sanka,R., Brinkac,L., Kariyawasam,S., Shelton,D., Fratamico,P.M., and Kapur,V. (2016). Whole genome sequencing of diverse Shiga toxin-producing and non-producing *Escherichia coli* strains reveals a variety of virulence and novel antibiotic resistance plasmids. Plasmid *83*, 8-11.

Qin,C., Yu,C., Shen,Y., Fang,X., Chen,L., Min,J., Cheng,J., Zhao,S., Xu,M., and Luo,Y. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proceedings of the National Academy of Sciences *111*, 5135-5140.

Schmid,D., Allerberger,F., Huhulescu,S., Pietzka,A., Amar,C., Kleta,S., Prager,R., Preussel,K., Aichinger,E., and Mellmann,A. (2014). Whole genome sequencing as a tool to investigate a cluster of seven cases of listeriosis in Austria and Germany, 2011-2013. Clinical Microbiology and Infection *20*, 431-436.

Stasiewicz,M.J., Oliver,H.F., Wiedmann,M., and den Bakker,H.C. (2015). Whole-genome sequencing allows for improved identification of persistent *Listeria monocytogenes* in food-associated environments. Applied and Environmental Microbiology *81*, 6024-6037.

Zhang,S., Yin,Y., Jones,M.B., Zhang,Z., Kaiser,B.L.D., Dinsmore,B.A., Fitzgerald,C., Fields,P.I., and Deng,X. (2015). *Salmonella* serotype determination utilizing high-throughput genome sequencing data. Journal of Clinical Microbiology *53*, 1685-1692.

Contact:

Greg Jones
Greg.jones@campdenbri.co.uk
+44(0)1386 842143