FINAL REPORT

ASSESSING THE POTENTIAL OF NOVEL MOLECULAR MICROBIOLOGICAL APPROACHES FOR MANAGING FOODBORNE DISEASE OUTBREAKS

FS101029

August 2012

CAROLE FOY AND DENISE O'SULLIVAN (LGC)

GREG JONES, SUZANNE JORDAN AND ROY BETTS (CAMPDEN BRI)

SARAH O'BRIEN (UNIVERSITY OF LIVERPOOL)

Prepared for the Food Standards Agency

Table of Contents

1. Executive Summary

Outbreaks of foodborne disease can cause serious public health consequences and significant political and trade issues. Current methodology for the investigation of foodborne disease outbreaks tends to be based on traditional microbiological techniques which are slow, have limitations and can result in delays which hamper identification of the source.

From a public health, policy and regulatory perspective, it is imperative to exploit any advances in knowledge and understanding of emerging technological developments which may assist control of foodborne disease outbreaks. Alternative molecular approaches such as PCR have been around for many years and have the potential to reduce timescales and improve investigation outcomes. However, such techniques are often only used in outbreak situations as a secondary investigative or confirmatory tool once traditional approaches such as culture have been conducted. The reasons for this are complex and include cost, validation considerations, concerns around variability/robustness, comparability to traditional approaches and "regulatory" acceptance of non-reference methods.

The field of genomics is currently evolving rapidly with new generations of technologies, such as ultra-high throughput sequencing, being developed that may provide improved knowledge and reduce outbreak investigation timescales further. This review outlines current and emerging genomic approaches that may have the potential to improve the management of foodborne disease outbreaks and provides examples of specific stages in the outbreak investigation process which would benefit from the use of molecular approaches. The review also highlights current capability and expertise in the UK and discusses routes and barriers to implementation.

In conclusion, these newer molecular approaches have the potential to significantly improve outbreak investigations and their use as a frontline tool is plausible. However, there remains a need for parallel development of standards and method validation approaches to determine accuracy, comparability, robustness and "fitness for purpose." Harmonised validation of these new "alternative" methods across the different laboratories providing data in to outbreak investigations is a critical requirement alongside the need for inter-operable databases. To facilitate adoption of molecular approaches robust datasets need to be generated to demonstrate "fitness for purpose" and a review of current practice should be undertaken to harmonise approval of novel tests in line with industrial guidelines and help reduce the burden on individual laboratories to implement and validate alternative technologies.

2. Disclaimer

Certain commercial equipment, instruments and assays are identified in this review by way of example. These examples are given to foster understanding only with other techniques and approaches available. Identification of specific instruments/assays does not imply recommendation or endorsement by the authors of this review, or that they are necessarily the best available for the purpose.

All trade marks are the property of their respective owners.

Certain information regarding instruments and assays has been obtained from publicly available sources. Genomics is fast a moving field, so figures quoted (costs, throughputs, installations etc) are a “snapshot” at a specific point in time only.

3. Glossary /Definitions

ACMSF	Advisory Committee on Microbiological Safety of Food
AFNOR	Association Française de Normalisation
Amplicon	Amplified template region produced during PCR
AOAC	Association of Analytical Communities
BIPM	Bureau International des Poid et Mesures
BSI	British Standards Institute
CE	Capillary Electrophoresis
cfu	Colony forming unit
Ct (also known as Cq)	Threshold (or quantification) cycle. The point in a real-time PCR amplification that signal can be detected above background
EmPCR	Emulsion PCR
fAFLP	Fluorescent amplified fragment length polymorphism
GI	Gastrointestinal
GSC	Genomics Standards Consortium
HMP	Human Microbiome Project
HPA	Health Protection Agency
HTA	Health Technology Assessment
HUS	haemolytic-uraemic syndrome
IID2	Infectious Intestinal Disease (IID) 2 study
IMS	Immuno magnetic separation
ISO	International Standards Organisation

MLST	Multi Locus Sequence Typing
Molecular	Methods based on examination of an aspect of an organism's Nucleic Acid
NGS	Next Generation Sequencing
OCT	Outbreak Control Team
OTU	Operational taxonomy unit
PCR	Polymerase Chain Reaction
POC	Point of care
PFGE	Pulsed field gel electrophoresis
pfu	Plaque forming unit
PT	Proficiency testing
qPCR	Real-time quantitative PCR
RFLP	Restriction fragment length polymorphism
rRNA	Ribosomal RNA
SGM	Society for General Microbiology
SNP	Single nucleotide polymorphism
VNTR	Variable number tandem repeat
VTEC (also known as STEC)	Vero cytotoxin-producing *E. coli* (Shiga toxin producing *E.coli)*

4. Background

This project was funded under the Food Standards Agency Strategic Challenges on Food Security call (FS246004 10/08/11), to address strategic challenge two: “How to use novel methods in managing foodborne disease outbreaks”.

A desk-based study was undertaken to identify and review the potential of emerging molecular microbiology technologies in the management of foodborne disease outbreaks. The objectives of the project were to:

1. Review current and emerging genomics technologies for rapid and accurate identification of causes of foodborne disease outbreaks
2. Review the potential of the technologies for improving current epidemiological practice
3. Review current expertise in the development and application of molecular approaches for managing outbreaks

5. Introduction

Accurate clinical diagnosis and rapid tracing of sources of foodborne disease outbreaks are vital for public health protection.

The term ‘outbreak’ is frequently applied in one of the following situations:

1. two or more related (i.e. epidemiologically linked) cases of a similar disease.
2. an increase in the observed incidence of cases over the expected incidence within a given time period.
3. a single case of a very serious disease like botulism.

Outbreak investigation needs to be systematic, thorough and rapid. Current methodology tends to be based on traditional microbiological techniques, such as culture, which can take days or even weeks for difficult-to-culture organisms. This results in avoidable delays and hampers identification of the source.

Outbreaks may come to light in several ways. Commonly they may be associated with particular premises and such point source outbreaks can be relatively easy to detect. More difficult are the outbreaks that are linked to a nationally or internationally distributed contaminated food product. These can result in small numbers of cases over a very wide geographical area, which can be more difficult to link together quickly. The longer it takes to establish a link the more difficult it can be to pinpoint a source, especially if the contaminated source is a short shelf-life perishable product which has disappeared from the distribution chain before an outbreak comes to light.

There are three strands of evidence in outbreak investigations that are woven together to implicate a source. These are:-

- microbiological evidence to (a) define the organism causing patients' symptoms and (b) to implicate a contaminated food vehicle by finding indistinguishable organisms in patients' clinical samples and in food samples;

- epidemiological evidence to implicate a food vehicle using statistical techniques, which can happen in advance of microbiological results being available. Indeed, epidemiological methods can be used to help focus in on the food vehicles to be tested, especially in a continuing source outbreak;

- environmental evidence to identify and explain the circumstances in which contamination of food vehicle took place. This might include microbiological sampling of the environment e.g. kitchen environment or the natural environment on an open farm.

In practice, these investigations run in parallel and the evidence from each is considered at regular intervals by an Outbreak Control Team (OCT). Depending upon circumstances some or all of these pieces of evidence are available to implicate a source. Actionable evidence (in law) tends to be microbiological.

It may be very difficult to identify a food vehicle in a continuing source outbreak for several reasons. Outbreak investigation often relies on asking ill people (cases) and well people (controls) about what they have eaten before the onset of symptoms in the cases. First, although people might be able to remember what meals they have eaten they might find it difficult to recall, or not even know, what ingredients were used in the meals that they ate. This can make it very difficult to pinpoint a contaminated ingredient. Secondly, people rarely blame themselves for foodborne illness. There is a tendency to identify the last meal eaten outside the home as the meal responsible for symptoms, which might not be the case. Given that foodborne illness is caused by a variety of agents (intoxination with an "incubation" period of 30 minutes to cryptosporidiosis with an incubation period of around three weeks) people who blame the last meal they ate outside the home will often be wrong. Thirdly, where a food vehicle is implicated statistically, there may be several explanations including:-

- *the implicated food was the cause of the patients' symptoms:* there are numerous examples in the literature where an epidemiological investigation correctly identifies a putative food vehicle, which is subsequently confirmed microbiologically (O'Brien 2012).
- *the implicated food item was not the cause of the symptoms but people who ate that food also ate another food that was the source of contamination:* this happened in an outbreak of Vero cytotoxin-producing *Escherichia coli* O104 in Germany when a high percentage of people who ate sprouted seeds also ate cucumber (Buchholz, Bernard et al. 2011), which was initially implicated as the source of the outbreak. Identification of a Vero cytotoxin-producing *Escherichia coli* (VTEC) led to a product withdrawal and sales of Spanish cucumber slumped. As events unfolded it became apparent that the cucumbers were contaminated with VTEC O8:H19, i.e. not the outbreak strain. Had sub-typing results been available earlier this might have alerted the OCT quickly that the implicated product, i.e. cucumber, was not the source of the patients' illness. Nevertheless, in the political furore that ensued people appeared to lose sight of the fact that the cucumbers should not have been contaminated with a VTEC in the first place.

- *a statistical association occurred by chance:* the greater the number of food items examined the greater the possibility of this happening although this should be minimised by careful attention to study design and statistical power.

Incorrectly identifying a source causes problems. First, the outbreak may continue to escalate as the real source continues to be consumed. Secondly, a food that is not contaminated might inadvertently be withdrawn from the market causing huge financial implications for producers and Governments. The incorrect citing of cucumbers as the source of the VTEC O104 outbreak has been reported to have costs of around £175 million per week, although, as pointed out above the cucumbers should not have been contaminated with a VTEC in the first place. Indeed, it is possible that an outbreak of VTEC O8:H19 was averted. In an outbreak of *Salmonella* Saintpaul in the US tomatoes were initially implicated before jalapeño and serrano peppers were found to be contaminated with the outbreak strain (Mody, Greene et al. 2011). This outbreak was reported to have cost the US tomato industry over US$100M (Thompson 2008). If a more rapid and accurate sub-typing system had been available, this might have been avoided, resulting in faster sourcing of the correct food vehicle.

In line with this, the Society for General Microbiology (SGM) produced a report in November 2011 entitled "Position statement on food security and safety" (http://www.sgm.ac.uk/PA_Forms/FoodPS_Web.pdf). Key research priorities identified included:

- Development of improved methods to detect food-borne microbes…including rapid detection of viable or infectious agents, for example in non-culturable states and to trace pathogenic and spoilage microbes through the food chain

- Development of rapid diagnostics to allow for correct [trace] species and strain identification from within mixed microbial communities

Ideally, front-line methods which are used to investigate foodborne disease outbreaks should be rapid (even if this includes a pre-enrichment step), accurate, affordable, sensitive, robust and informative. Detecting and confirming the presence of the causative organism rapidly and accurately may require the detection and identification of multiple pathogens in one test while simultaneously identifying virulence or pathogenicity determinants. Additionally, outbreaks may need to be further characterised by sub-typing or other methods, to pinpoint casual links as demonstrated by the VTEC O104:H4 and salmonellosis outbreaks described above. Currently, if a sample (food or patient) needs to be screened for unknown agents, then many separate tests have to be conducted. If methods could be devised and validated which simultaneously allowed detection of multiple agents to Genera/Species/Strains level from uncultured samples, then test time and laboratory time could be significantly reduced and results obtained more quickly and more cost effectively.

Current molecular approaches such as PCR offer many advantages over traditional methods as the sensitivity and specificity of the approach means that they do not usually have to rely on an initial culture step. This enables the rapid detection of

organisms including those that are either difficult to culture or cannot be cultured in a laboratory.

However, molecular approaches raise challenges in terms of correlation with traditional methods. Culture approaches will only detect viable organisms that can be cultured in a laboratory. For quantitative assays the reported value is usually colony forming units (cfu) or plaque forming units (pfu) which will be dependant upon culture conditions used and is not directly equivalent to numbers of organisms present (i.e one cfu is not always equal to one bacterium). In contrast, PCR based approaches will detect the genetic material of viable, non-viable and sub-lethally damaged organisms with the reported value being either arbitrary units (such as PCR threshold cycle) or copies of a genome. Consequently, the differing sensitivities/specificities of the approaches, the differing properties being measured and the differing measurement units make comparisons between approaches difficult and the validation of "alternative" methods challenging.

Molecular approaches are also considered expensive and consequently are not generally used early in the outbreak process. A review published by the Health Technology Assessment (HTA) Programme in September 2007 (Abubakar, Irvine et al. 2007) assessed the potential of available molecular approaches at that time. The aim of the review was: "To determine the diagnostic accuracy of tests for the rapid diagnosis of bacterial food poisoning in clinical and public health practice and to estimate the cost-effectiveness of these assays in a hypothetical population in order to inform policy on the use of these tests".

The review of rapid diagnostic tests for six organisms selected examined 70 studies focussed on food, and 87 clinical studies. The conclusions were that despite the promising correlation in diagnostic accuracy between rapid methods and conventional culture techniques, there were few studies that made this comparison directly. Therefore, the larger proportion of positive results obtained from the rapid methods studied was difficult to interpret. Whether the additional positive samples were due to a lack of sensitivity of culture based methods, or due to a lack of specificity of the rapid methods was difficult to ascertain.

The overall conclusion from the report was that despite the clear advantages of nucleic acid based techniques in terms of decreased time from sample submission to results, there were still factors that prevented their more extensive use in clinical and food microbiology settings. In particular, the paucity of data regarding the increased number of positive samples found with nucleic acid analysis, and the implications of this, would be a major obstacle to the adoption of these methods.

The last decade, and particularly the last five years since the 2007 HTA report, has seen an increase in the number of higher throughput and/or more extensive analytical technologies that offer unique opportunities for molecular epidemiology. During this period the cost of performing many of these types of assays has continued to reduce, making them a more attractive proposition. Additional studies investigating the validity of data generated by molecular platforms have also been undertaken.

Genomic approaches such as next generation sequencing (NGS), high-throughput microfluidic PCR, multiplex and real-time PCR, mass spectrometry and microarrays are being developed which are capable of accurately and sensitively detecting,

subtyping, characterising and quantifying existing and new/emerging pathogens. The rapid sequencing of the genomes of disease-causing organisms and subsequent development of molecular diagnostics is now occurring. For example, in the 2011 VTEC O104:H4 outbreak in Germany, NGS was used to sequence the pathogen's genome in just three days. This was quickly followed by the development of a quantitative PCR (qPCR) based diagnostic for the new strain (http://www.iontorrent.com/lib/images/PDFs/co23298_ecoli.pdf). In parallel with the emergence of newer generations of high throughput genomics approaches, there have also been significant developments in the field of rapid, portable, disposable test systems capable of detecting pathogens at the "point of care" or point of sampling.

These newer, higher throughput and more informative molecular approaches are currently revolutionising molecular biology and may offer opportunities for improved management of outbreaks if used earlier in the process. These powerful molecular approaches could be employed to start to identify subtype and virulence or pathogenicity determinants in outbreak situations, to attribute cause quickly and correctly. These methods are very important tools as, only by identifying true cause, can an outbreak be stopped.

However, as discussed above, whilst these emerging approaches offer much, they also present challenges when considering implementation in current laboratory systems, and comparability and standardisation of measurement. Harmonised validation of these new "alternative" methods is a critical requirement. Currently, methods used by food microbiologists must be either recognised as reference methods eg. ISO, AOAC, or validated to demonstrate at least equal performance to reference methods. As discussed above, this poses very significant challenges as not only are there differences in the measurement approaches, but also what is measured is different (eg. DNA sequence) and not necessarily defined by the growth media, or conventional culture conditions.

The following sections will review current, new and emerging molecular microbiological methods that could be employed within the epidemiological process for identifying and tracing the source of a food poisoning outbreak. Later sections will highlight some of the challenges associated with implementation of newer technologies.

6. Search strategy

A strategy for gathering information for this review was developed. Relevant peer-reviewed articles and reviews were identified through searching the electronic database PubMed using combinations of terms including; molecular methods, molecular epidemiology, outbreak, identification, detection, foodborne disease, foodborne pathogens, emerging methods, diagnosis, sequencing, PCR, genomics. These terms were also used in the web program Google Scholar to overcome the predominantly biomedical focus of the PubMed database. Database searches were primarily limited to publications post September 2007.

Websites from leading manufacturers of molecular tools and instrumentation were examined to identify assays or instruments that could be used in food microbiology and molecular epidemiology. From this relevant manufacturers were contacted for further information about the technology. Market reports and trade journals were examined for relevant information. Conference proceedings, workshop reports and poster abstracts were included in the search for relevant material. Experts in the field identified from personal recommendation, publications and workshop reports etc. were contacted for advice.

Annual reports from the HPA on Gastrointestinal (GI) infections and user manuals for the HPA Laboratory of Gastointestinal Pathogens were also consulted. In addition, the 2007 Health Technology Assessment (HTA) report entitled “A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food” was a useful starting point for this review. The HTA report summarised rapid methods and techniques developed up to and including September 2007.

7. Overview of current and emerging molecular testing approaches

The principles of currently widely used and well-established molecular approaches such as the Polymerase Chain Reaction (PCR) and sequencing will not be described in detail in this report. Appendix 1 provides references and a brief description of the fundamentals of these approaches and other well-established molecular approaches used by molecular microbiologists including Pulsed Field Gel Electrophoresis (PFGE), ribotyping, Variable Number Tandem Repeat (VNTR) profiling, fluorescent amplified-fragment length polymorphism (fALFP) and Multilocus Sequence Typing (MLST). Whilst newer PCR and sequencing technologies are largely improving or replacing these older techniques in terms of technical performance, the older techniques are still widely used due to the availability of reference databases containing profiles generated using those approaches. Table 1 briefly summarises the main features of the currently used molecular methods in terms of factors such as ease of use, cost and throughput.

Method	Pure Culture? (Y/N)[1]	Cost[2]	Time to result[3]	Ease of use[4]	Ease of automation	Throughput	Reference databases available? (Y/N)	Validated 3rd party assays available? (Y/N)	Identification level	Where it is currently used.
PFGE	Y	Med	4 to 5 days	Difficult	Difficult	Low	Y	N	Sub species/serotype	Clinical & Food
Ribotyping	Y	Med	2 to 5 days	Easy	Easy	Low	Y	Y	Sub species/serotype	Food
MLST	Y	High	2 to 5 days	Difficult	Difficult	Low	Y	N	Sub species/serotype	Clinical
VNTR	Y	Med	2 to 5 days	Medium	Medium	Low	Y	N	Sub species/serotype	Clinical & Food
fAFLP	Y	Med	2 to 5 days	Medium	Difficult	Low	N	N	Sub species/serotype	Clinical & Food
Standard PCR[7]	N	Low	24 to 48 hours	Easy	Medium	Medium	Y	Y	Detection of Genus or species [5]. Typing methods at sub species level also available	Clinical & Food
Real-Time PCR	N	Low	24 to 48 hours	Easy	Easy	High	Y	Y	Detection of Genus or species [5]	Clinical & Food
Sanger Sequencing	Y	Med	2 to 5 days	Difficult	Medium	Low, will depend upon the instrument used	Y	Y	Generally species/sub species[6]	Clinical & Food

[1]Does the method require a pure culture of the target organism before it can be used?

[2]Costs are approximate, and will depend on an individual laboratory's labour charges etc. Low: <£100 per sample, Medium: £100 - £500 per sample, High: >£500 per sample. Costs do not consider any capital equipment costs, only per sample consumable costs.

[3]Time from initial sample submission to result. Different samples and methods will result in slightly longer or shorter times to result, so the figures here are indications only.

[4]The ease of use reflects the level of training for staff using the technique

[5] Dependant upon the detection target of primers and probes

[6] Dependant upon the information available for the organism concerned

[7] Not generally used in the food industry for pathogen detection as most assays have been adapted to real time systems to reduce issues with sample cross contamination through the use of a closed system

TABLE 1 SUMMARY OF CURRENT MOLECULAR APPROACHES

The following sections of this review will focus on more recent developments in molecular technologies. The following flow chart summarises some of the major developments in molecular technologies over the past 10-20 years with ongoing refinements and improvements predicted in to the future.

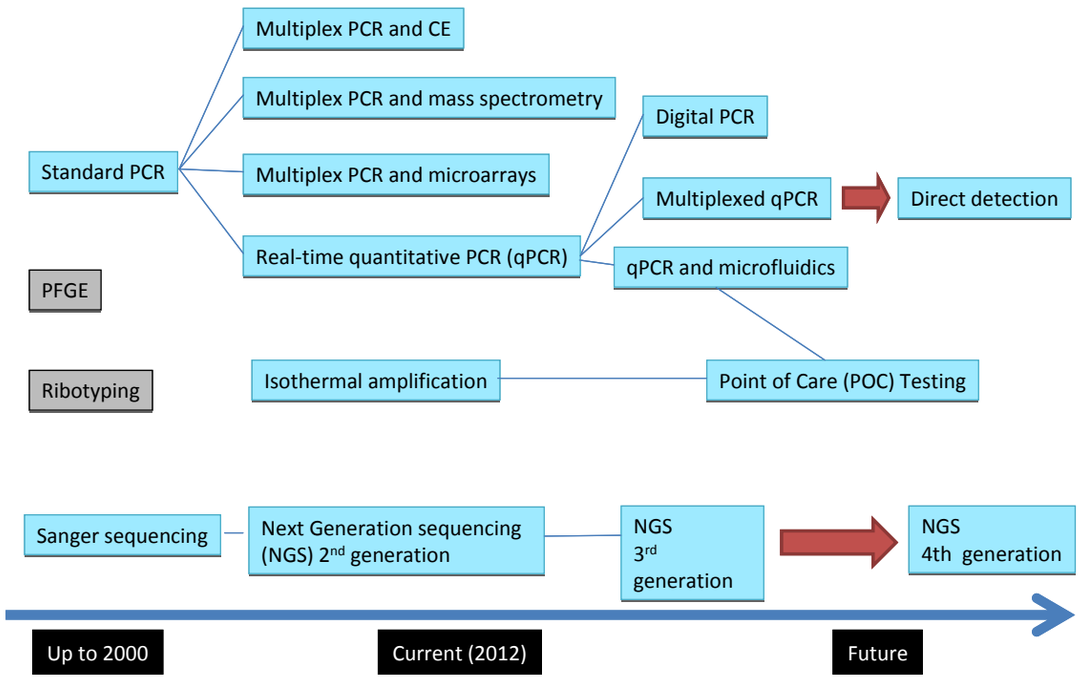


FIGURE 1 CURRENT AND EMERGING MOLECULAR APPROACHES

The following sections will describe the novel and emerging approaches outlined above in more detail; discuss where they are currently being used in food outbreak situations and where they further have the potential to add value.

7.1. PCR based developments

The polymerase chain reaction (PCR) was developed in the early 1980s by Mullis and co-workers and used an enzyme (DNA polymerase) together with oligonucleotide primers specific for target DNA regions to amplify that target region

over a billion fold. (Mullis K 1986) Since then, PCR has become established as the dominant nucleic acid analysis technology across a wide variety of sectors with a plethora of PCR approaches introduced in both clinical and non-clinical settings to attempt to speed up the pathogen identification process.

Whilst many laboratories design and develop their own “in-house” molecular assays there are also many commercial PCR based systems for detecting foodborne organisms. Figure 2 shows the number of kits available commercially for use in the food industry. These data are taken from a recent review by Campden BRI entitled “Catalogue of rapid microbiological methods” (Campden BRI 2012).

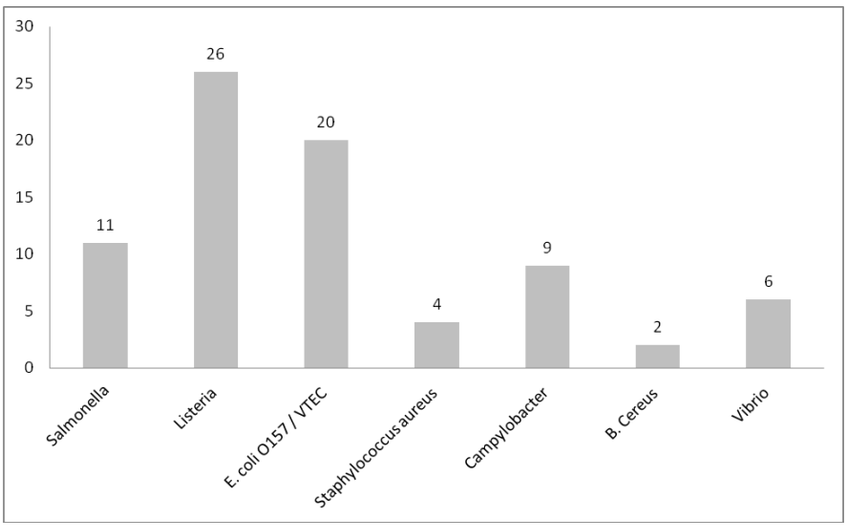


FIGURE 2 NUMBER OF PCR BASED KITS (Y-AXIS) AVAILABLE TO THE FOOD INDUSTRY DISPLAYED ‘PER ORGANISM’ (X-AXIS)

Whilst these kits have a broadly similar mode of operation (enrichment, followed by a PCR reaction), some claim faster time from initial sample submission to identification. Appendix 2 lists examples of kits available, the type of PCR used, their claimed time to identification and their status with various validation bodies.

On-going improvements to the basic PCR process have included increasing the throughput of the approach by a variety of multiplexing and downstream analysis strategies and improving the sensitivity and quantitative potential of the technology through a variety of enrichment and real-time PCR strategies. The following diagram outlines the four major stages in the PCR process, namely sampling, extraction of nucleic acid, amplification and detection of amplified product (amplicon). The major options currently available for the detection stage are also highlighted.

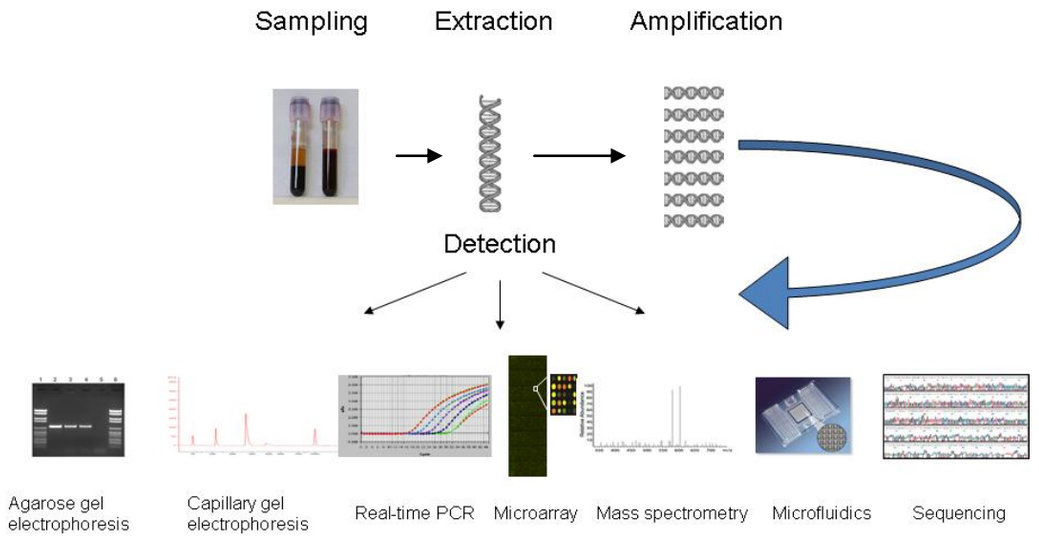


FIGURE 3 OVERVIEW OF CURRENT PCR BASED APPROACHES

PCR based technologies have been recognised by British Standards and three standard methods that provide guidance on the sample preparation and requirements for the qualitative detection of food borne pathogens by PCR have been issued (BS EN22174: 2005, BS EN 20837:2006 and BS EN 20838:2006). These standards define the criteria to be used when designing and carrying out the assay rather than being prescriptive in the method details. The standards require the inclusion of positive and negative controls for the extraction and PCR amplification steps. A further control is also needed to detect inhibitory compounds in the sample or enrichment broth that may result in a false negative result being obtained.

The adoption of PCR technologies for the detection of foodborne pathogens as reference methods is becoming a reality with two draft protocols in preparation for toxigenic strains of *Clostridium botulinum* and selected VTEC serotypes. VTEC are a challenging group of organisms to detect in food samples due to their potential presence in low numbers, their sensitivity to potential selective agents and the availability of limited tests that enable their discrimination from closely related organisms, which lends itself well to molecular based assays. The proposed draft method for 5 defined serotypes of VTEC detects members of this highly pathogenic group in two distinct steps. The first step uses target genes that define its characteristics: the ability to produce shiga toxins through the presence of stx1 and/or stx 2 as well as the ability to attach to the GI tract using intimin encoded by eae. Samples that contain the genes characteristic of VTEC are taken forward to the second step that differentiates between the 5 VTEC serotypes of interest using the genes encoding the lipopolysaccharide O factors specific to each serotype. This approach affords flexibility to expand the assay to include other serotypes that are highlighted to be of importance in the future such as the O104:H4 which has been added to the serotypes listed in the latest EU criteria under consideration for sprouted seeds.

The following sections will outline some of the latest PCR based approaches currently available in more detail.

7.1.1. Real-time quantitative PCR (qPCR)

Real-time quantitative PCR methods (qPCR) typically use primers (as for standard PCR) in conjunction with fluorescent dyes that bind to the amplicon produced during PCR or fluorescently labelled oligonucleotide probes that hybridise to a specific region within the PCR amplicon for increased specificity. As the amplicon accumulates during the PCR process a fluorescence read-out is taken at each cycle of the amplification process. The amount of signal detected and the point in the amplification process where signal can be detected above background (typically called the threshold cycle (Ct) or quantification cycle (Cq)) correlates to the amount of target present in the starting sample. Therefore, accurate quantification of the starting amount of target present is possible.

Advantages of this technique are that quantification is possible and the process is easily automated with no need to re-open tubes post-PCR thereby reducing the risk of contamination seen with standard PCR. The ability to detect the product during the PCR amplification process also reduces assay time as additional post-PCR detection procedures are not required. A digital recording of the data also has advantages for ensuring traceability and data comparability.

Disadvantages of the technology are that no amplicon size confirmation is possible and it is currently challenging to multiplex more than 3-4 targets together due to instrument and fluorescent dye limitations. Data analysis can also be challenging with a need to set appropriate thresholds and cut-offs to define positive amplification.

Real-time PCR approaches are rapidly replacing standard PCR approaches due to their ease of use and quantitative potential and many different commercial kits are available for the detection of foodborne pathogens (see appendix 2).

7.1.2. Multiplex PCR

Many of the recent developments in the PCR format have involved the amplification of multiple targets within a single tube (multiplex PCR) followed by a variety of downstream approaches for determining which targets have amplified.

These downstream approaches vary according to the number of targets being detected simultaneously and include:

- real-time PCR approaches for the detection of small numbers of targets (<6)
- capillary electrophoresis for the detection of around 20 targets
- microarray platforms using solid surface or bead/particle-based formats for the detection of medium to large numbers of targets (ranging from 20 targets to hundreds of thousands of targets depending on platform)
- discrimination of PCR amplicons based on unique mass using mass spectrometry for medium numbers of targets (around 50- to several hundred)

- discrimination of PCR amplicons based on sequencing approaches (to be discussed in later sections)

Advantages of multiplex PCR approaches include the parallel detection of numerous targets (from tens to tens-of-thousands) in one assay which can provide rapid characterisation of a sample while saving on cost and resources.

Disadvantages of multiplex PCR amplification include the potential lack of specificity due to mispriming events and the out-competing of some assays by other more efficient assays or high abundance targets. Data analysis can also be challenging. The isolation of sufficient material from the food sample can also pose problems when multiple targets are being interrogated. However, recent developments in whole genome amplification and pre-amplification strategies are helping to overcome some of these issues.

7.1.2.1. Multiplex qPCR

Fukushima *et al*. used a multiplex qPCR assay to detect 24 different target genes in 8 multiplex reactions in DNA extracted from stool samples following a foodborne outbreak (Fukushima, Kawase et al. 2010). The specific assays were able to detect 16 different bacterial species including: *E. coli*, *Shigella* spp., *Salmonella* spp., *Campylobacter jejuni*, *Staphylococcus aureus*, *Clostridium perfringens* and *Listeria monocytogenes*.. The authors reported that the method was able to detect the causative agent in 33 outbreaks of foodborne disease in 3 hours or less and could potentially detect more (true) positives than are currently being reported from culture results alone.

Huang *et al*. used multiplex qPCR and multiple probes for the detection of multiple target sequences for species identification within the one reaction (Huang, Hu et al. 2007). The assays targeted 8 foodborne pathogens including *S. aureus*, *L. monocytogene*, Shigella spp. and *E. coli* O157:H7 and was assessed on both clinical and food isolates. Concordance with traditional methods was seen.

Real-time PCR approaches have also been widely used across other areas of clinical microbiology. For example, Lehmann *et al* (Lehmann, Hunfeld et al. 2008) used this approach for the early detection of bloodstream infections in septic patients including *E.coli*. A multiplex real-time PCR-based assay for the detection of 25 clinically important pathogens directly from whole blood in less than 6 hours was described. Comparing PCR identification results with conventional microbiology for 1,548 clinical isolates yielded an overall specificity of 98.8%. The analytical specificity in 102 healthy blood donors was 100%. The authors concluded that the assay holds promise for more rapid pathogen identification.

7.1.2.2. Capillary electrophoresis

Standard PCR typically uses agarose gel electrophoresis for visualisation of PCR amplicons and identification based on amplicon size. Capillary electrophoresis offers greater spatial resolution than agarose gels and instruments are available that can detect amplicons labelled with different coloured fluorescent dyes simultaneously,

thereby increasing the multiplex potential. Leader *et al* (Leader, Frye et al. 2009) combined multiplex PCR and capillary electrophoresis to enable high-throughput molecular determination of *Salmonella enterica* serovars based on fragment size separation. Sixteen genomic targets were interrogated based on their differential distribution among common serovars and 751 clinical isolates were assessed that had previously been serotyped via antisera. In total 89.6% of the isolates were correctly identified. The authors were of the opinion that this approach allowed simple and accurate typing of the most prevalent clinical serovars of Salmonella enterica at a level comparable to that of conventional serotyping, but at a fraction of both the cost and time required per test.

7.1.2.3. Microarrays

Identification of foodborne pathogens can be undertaken using microarray technology. Arrays typically consist of a surface, usually a glass slide, chip or bead, to which selected oligonucleotide probes are bound at specific locations. Fluorescently labelled DNA (or PCR amplicon) from the sample of interest is hybridised to the array where it binds to the region of the array containing a complementary probe resulting in fluorescence signal at that location. Alternative signal generation chemistries such as colorimetric or chemiluminescent outputs have also been developed.

Advances in microfabrication and array spotting technologies have resulted in the development of high density arrays with over one million discrete probe locations possible per array. Lower density arrays (hundreds of probes) are typically used for the detection of foodborne pathogens to facilitate rapid and cost-effective screening of larger sample numbers.

The microarray method requires purification of the starting material and labelling with fluorescent dyes (can be incorporated during PCR). The probes on the chip can cover specific genetic fragments of interest or the whole genetic sequence of an organism. The chip containing the hybridised sample of interest is washed to eliminate non-specific binding events, and then scanned and the fluorescence of the various spots (locations) on the chip quantified. The higher density arrays (several thousand spots) typically require a specialised array scanner and specialised software for analysis of the data.

A commercially available example of an array platform used for detecting foodborne pathogens is the Identibac *E. coli* Genotyping array from Alere Technologies (Alere Technologies GmbH). This array format consists of 146 probes printed onto an array located in the bottom of an ArrayTube (Figure 4) for the detection of virulence genes in *Escherichia coli* isolates. The probes represent 92 virulence genes but to encompass gene variability some genes are represented by more than one probe.

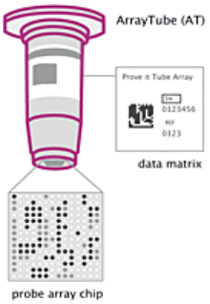


FIGURE 4 ARRAYTUBE ASSAY (obtained from http://alere-technologies.com/)

Geue *et al*, reported use of the above technology for the genotyping of enterohemorrhagic *Escherichia coli* serotype O156:H25/H−/Hnt Isolates from Cattle (Geue, Schares et al. 2010) and Anjum *et al* also used this approach to pathotype *E.coli* strains (Anjum, Mafura et al. 2007).

A commercially available example of a bead array is the Luminex xTAG System. The Luminex platform is based on the principle of flow cytometry. The system enables the simultaneous multiplexing of up to 100 analytes in a single microplate well. The system uses up to 100 differentially fluorescently labelled polystyrene microspheres that act as both the identifier and the solid surface to build the assay. The process takes approximately 5 hours and incorporates extraction and amplification of DNA, hybridisation of the PCR product to beads on a suspension array platform, followed by acquisition and analysis of data.

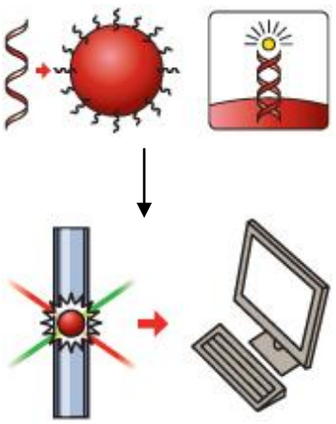


FIGURE 5 OVERVIEW OF THE LUMINEX XTAG SYSTEM (obtained from http://www.luminexcorp.com/) PCR products are hybridised to a panel of up to 100 different beadsets and then analysed by flow cytometry

Luminex's xTAG Gastrointestinal Pathogen Panel (GPP) offers a qualitative multiplex test which can identify 15 of the major gastrointestinal pathogens (9 bacterial, 3 viral and 3 parasitic) that are responsible for over 95% of cases of infectious diarrhoea. The assay is indicated for patients with signs and symptoms of infectious gastroenteritis or colitis, acute and chronic diarrhoea, inpatient, outpatient and emergency settings and nosocomial surveillance.

Malecki *et al.* reported the use of the Luminex system at the time of the VTEC *E. coli* O104:H4 outbreak in Germany for rapidly screening patients suffering from haemorrhagic diarrhoea and at the risk of developing the serious and life-threatening haemolytic-uremic syndrome (Malecki, Schildgen et al. 2012). Such patients need a rapid differential diagnosis to avoid misdiagnosis potentially leading to incorrect initial therapy or infection control measures. A total of 20 patients were tested with 4 patients testing positive for the new O104:H4 strain, of which 2 had previously been independently confirmed by an external laboratory. The authors concluded that the assay is useful for prescreening patients and is suitable for high throughput analyses to cover the peaks in an epidemiologic outbreak situation.

Tankouo-Sandjong *et al.* developed an oligonucleotide microarray method for Salmonella serotyping (Tankouo-Sandjong, Sessitsch et al. 2008), and Fitzgerald *et al*. used a multiplex, bead-based suspension array for the molecular determination of six common Salmonella serogroups (Fitzgerald, Collins et al. 2007).

Bai *et al*. used multiplex PCR and optical biosensor chips to detect 11 pathogens likely to cause foodborne disease on DNA extracted from culture and found the technique to be sensitive, with a detection limit of 85 cfu/mL, specific and rapid with the assay taking 30 minutes to complete after the PCR reaction had been completed (Bai, Zhao et al. 2010).

7.1.2.4. Mass spectrometry

PCR coupled with mass spectrometry has been described to detect foodborne bacteria in samples (Mazzeo, Sorrentino et al. 2006; Boxrud 2010). These spectrometric techniques typically ionize DNA or PCR amplicons from clinical or food isolates. The ions are accelerated and transferred to an analyzer where they are separated and detected based on the ratio of molecular weight to charge (m/z). The mass spectra obtained can be considered specific fingerprints or molecular profiles of the bacteria analyzed and can also detect mixed populations of microbes simultaneously within a sample.

The key requirement is the linking of the Mass Spectroscopic system with a validated database of spectra from known named organisms. This enables rapid comparison of spectra from unknown samples with the database and potentially identification within a few minutes. The published literature on microbial identification using mass spectrometry based systems is beginning to rapidly expand and on the whole indicates a very favourable comparison to conventional biochemical identification methods with a reduction in the per sample cost (Cherkaoui 2010.).

Mazzeo *et al* (Mazzeo, Sorrentino et al. 2006) described a methodology based on matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF) for the rapid discrimination of 24 bacterial species belonging to the genera *Escherichia, Yersinia, Proteus, Morganella*, and *Salmonella*.

PCR coupled with mass spectrometry has also been reported by Hannis *et al* who carried out high-resolution genotyping of *Campylobacter* species by the use of PCR and high-throughput mass spectrometry (Hannis, Manalili et al. 2008). The method

was automated and reported to have a resolving power comparable to that of multilocus sequence typing (MLST). Amplicons were derived from PCR primers which amplify short (<140-bp) regions of the genes used by conventional MLST strategies, making it potentially cross-comparable with existing databases. The reported sensitivity was 10 genomes/PCR reaction from pure isolates.

Commercial platforms based on mass spectrometry are also available. The PLEX-ID system from Abbott Ibis Biosciences involves PCR amplification of generic or specific microbial targets followed by identification of the PCR amplicons by mass using electrospray ionisation mass spectrometry. From a database, the mass of the amplicon can be used to infer the sequence composition which can then be used to determine the identity of the pathogen and characterise the sample. The system allows for screening for known and unknown bacteria, viruses, or fungi in a single analysis directly from the sample. Up to 250 samples per day can be analysed with assay turnaround times of approximately eight hours.

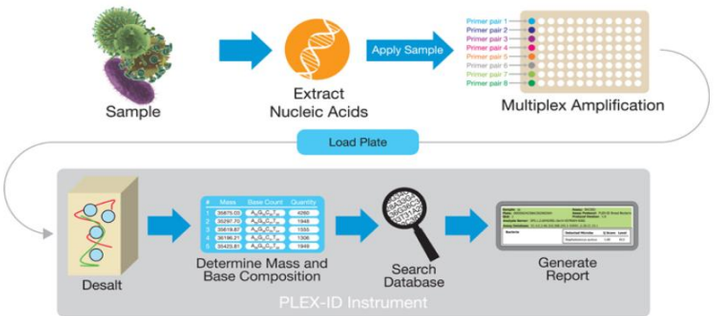
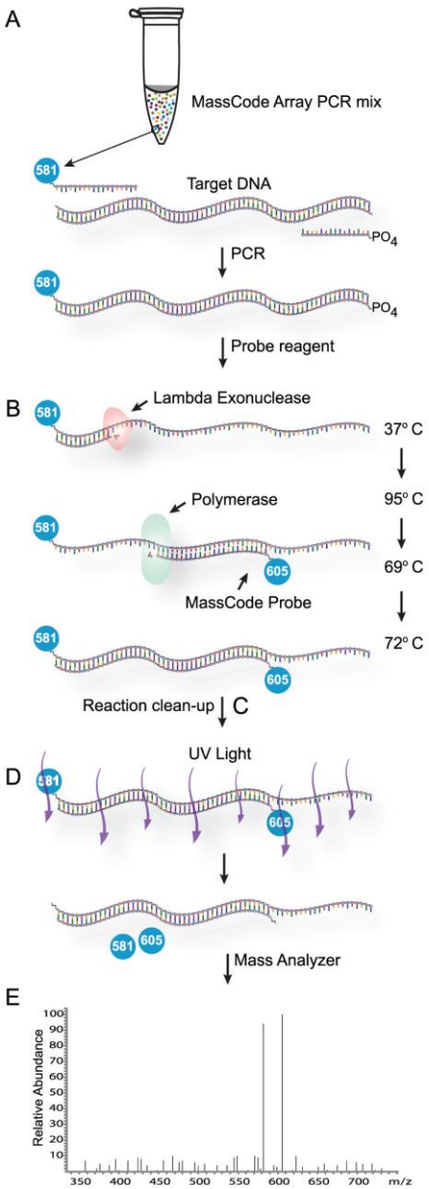


FIGURE 6 OVERVIEW OF THE PLEX-ID SYSTEM (obtained from http://www.ibisbiosciences.com/)

PLEX-ID assays are available in three forms: 1) broad assays to identify a wide range of bacteria, virus, or fungi. 2) targeted assays to identify a specific set of organisms, and 3) characterization assays which provide strain-or-drug resistance/virulence information for a specific species. The Broad Bacteria assay identifies more than 3,100 species of bacteria direct from isolates and the Food Borne Bacteria targeted assay can detect and identify several common food-borne bacteria including *E. coli* O157:H7 and non-O157 VTECs, over 250 members of *Salmonella* enterica species, *Shigella*, and *Listeria* species.

A related approach has been developed by Agilent based on MassCode(MC) Tags. These tags are small molecules, each with a unique molecular weight which are attached to DNA primers corresponding to specific microbial sequences. Following PCR amplification the tags are released by applying UV irradiation and mass spectroscopy is then used to identify the corresponding tags. The use of this technology and proof of principle demonstration of a 14-plex assay to subtype a select panel of Salmonella enterica serogroups and serovars was described by Richmond *et al*. (Richmond, Khine et al. 2011)

As with other mass spectrometry based systems further studies will need to be performed to validate these emerging approaches.



A) Workflow begins with a MC-PCR reaction containing all primer pairs for each target group.

B) MC-PCR products containing 5′ phosphate strands undergo digestion at 37°C with lambda exonuclease.

After digestion a second round of selection is performed. MassCode probes are annealed to internal sequence and serve as extension primers to result in a double-strand single-strand segmented hybrid labelled with two unique MassCode reporters.

C) Unincorporated oligonucleotides and misprimed amplified DNA less than 100 bp are removed during a reaction clean-up step.

D) and E) MCTs are cleaved from the hybrids upon exposure to UV light and flowed directly into a single quadrupole mass spectrometer for detection.

FIGURE 7 OVERVIEW OF THE MASSCODE TAG APPROACH (Obtained from:(Richmond, Khine et al. 2011))

7.1.3. High throughput qPCR and microfluidics

New innovations in microfluidics and miniaturisation of the PCR process have taken real-time qPCR platforms to a new level in terms of sample throughput. Scaling of the PCR reaction from the microlitre to the nanolitre range dramatically reduces the volumes of reagents and samples required and enables many thousands of assays to be conducted simultaneously.

High-throughput qPCR may be particularly beneficial for rapid screening of multiple pathogens. These approaches also overcome the issues associated with multiplex

PCR by allowing single assays to be performed in a highly parallel and high throughput manner.

However, to enable detection of target organisms in nanolitre volumes requires presence of the organisms in the starting material at sufficiently high concentrations. Strategies such as enrichment or pre-amplification of target material often have to be considered prior to the use of microfluidic qPCR systems.

Commercial examples of high throughput qPCR platforms include the Fluidigm BioMark system and the LifeTechnologies OpenArray system.

The BioMark system uses microfluidic chips (dynamic arrays) consisting of a network of capillary channels, valves and reaction chambers. Multiple chip formats are available with common formats consisting of 48x48 or 96x96 arrays. In the 48x48 format 48 assays can be loaded into the 48 assay inlets on one side of the chip and mixed in every pair-wise combination with 48 test samples loaded into the 48 sample inlets on the other side of the chip. Pair-wise mixing of samples and assays results in 2,304 individual amplifications in 6nl assay chambers on the chip. The 96x96 array format allows 9,216 individual assays to be performed. The platform can be used with multiple standard qPCR chemistries.

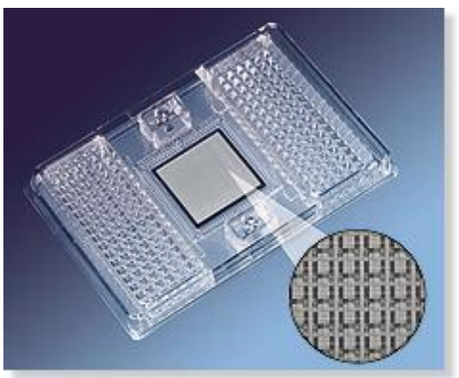


FIGURE 8 FLUIDIGM BIOMARK MICROFLUIDIC CHIP (obtained from http://www.fluidigm.com/)

Bugarel *et al* used this approach to identify genetic markers for differentiation of vero cytotoxin-producing, enteropathogenic, and avirulent strains of *E. coli* O26 (Bugarel, Beutin et al. 2011). They used the 48x48 microfluidic qPCR array format for the selection of discriminative genetic markers among 33 tested genes. They concluded that this approach might be applicable in hospital service laboratories or public health laboratories to test strains isolated from stools of patients suffering from diarrhoea.

The OpenArray platform (Life Technologies) comprises a microscope slide-sized stainless steel chip consisting of 48 sub-arrays each containing 64 (8×8) through-holes (reaction chambers), with a total of 3,072 qPCR chambers per array. For qPCR applications, assays are pre-spotted on the interior surface during the manufacturing process. The associated PCR cycler can process three chips

simultaneously. Both custom arrays and off-the-shelf target panels are available for screening samples.

FIGURE 9 OPENARRAY HIGH THROUGHPUT PCR PLATFORM (obtained from http://www.lifetechnologies.com)

Gonzales *et al*. used the OpenArray format to investigate 28 target genes involved in virulence, antigeniticy and regulation in one sample to identify, virulotype, and subtype O157 and non-O157 VTEC *E. coli*. They reported it to be a sensitive, specific and high throughput method with the potential to become an integral tool in outbreak, environmental, and genetic investigations of EHEC (Gonzales, Kulow et al. 2011).

7.1.4. PCR and point of care (POC)

Although PCR based methods are generally quicker than traditional approaches the samples still need to be transported to a specialised laboratory where nucleic acid from the sample is extracted, analysed and findings reported. More recent advances in molecular techniques, miniaturisation, microfluidics and instrumentation are leading to the emergence of near-patient or decentralised testing systems. These approaches have the potential to produce results in minutes rather than hours or days and could be used as an initial screen to rule particular pathogens in or out, followed by further characterisation in the laboratory.

Market providers include Cepheid with tests currently available for infectious diseases such as influenza and norovirus and healthcare acquired infections such as MRSA and *C.difficile*. Many more companies and research groups have products in development so we can expect to see an increase in products reaching market in the next few years.

FIGURE 10 GENE XPERT POC SYSTEM (obtained from http://www.cepheid.com/)

Further advances include the development of Isothermal technologies which, unlike PCR, involve amplification at a constant temperature. These approaches could provide a more cost effective POC screening test as they do not require thermocycling equipment.

7.2. Isothermal amplification based developments

Very recently a number of commercial method producers have released test systems for detecting foodborne pathogens that are based on isothermal amplification techniques. Such methods allow amplification of target nucleic acids at a single temperature, rather than the multiple temperature approach required in traditional PCR.

At present the most common system used commercially is Loop Mediated Isothermal Amplification (LAMP) which was first described in 2000 (Notomi T 2000) which can be used to amplify DNA or RNA, the latter a via reverse transcripase step.

The currently commercialised systems for detection of pathogens in foods based on LAMP techniques are reported to be simple to use and have received some validation showing equivalency to accepted reference methods for their target pathogens (personal communication).

Other isothermal amplification techniques have also been developed and used, including: Nucleic Acid Sequence Based Amplification (NASBA), Helicase-Dependent Amplification (HDA), Rolling Circle Amplification (RCA) and Strand Displacement Amplification (SDA). Full reviews of these techniques are available in published papers (Gill 2008) (Asiello 2011)

The main advantage of isothermal techniques over conventional PCR is that there is no need to cycle temperatures during the amplification reaction. This can make instrumentation smaller, more energy efficient, potentially lower in cost and more portable. Isothermal amplification reactions also tend to be faster than PCR based amplifications. These two factors mean that isothermal amplification methods may be particularly suited to POC applications.

7.3. Sequencing based developments

Methods for determining the precise sequence of a DNA molecule were developed in the 1970s and are considered the “gold standard” molecular approach for confirming the identity of a DNA target. However, until recently, the high cost and low throughput capability of sequencing approaches had restricted its use to limited sequencing of fairly short stretches of DNA or confirmation of identity of PCR amplicons. That situation started to change rapidly at the beginning of the 21st century when ground-breaking improvements to the throughput and concomitant reductions in cost began to occur.

7.3.1. Standard Sanger sequencing

Sanger sequencing was developed in the 1970’s (see appendix 1) and until relatively recently was the only practical way to sequence genetic material. The relatively low throughput and high cost of standard sequencing approaches compared to traditional microbiology means that this technique is typically used either as a confirmatory tool, or as a method of gleaning information that the traditional methods have been unable to provide.

7.3.2. Next Generation Sequencing – Second generation

Next generation sequencing (NGS), also known as massively parallel or ultra high-throughput sequencing describes a major shift from the “first generation” Sanger sequencing described above in terms of increased throughput and reduced costs. Figure 11 shows the dramatic reduction in costs from 2008 onwards following the advent of NGS technologies.

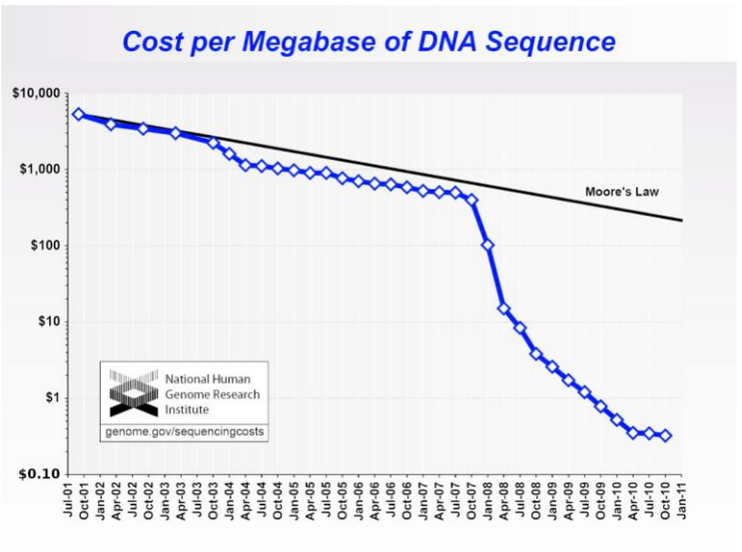


FIGURE 11 SEQUENCING COSTS PER MEGABASE OF DNA (obtained from wetterstrand ka. DNA sequencing costs: data from the nhgri large-scale genome sequencing program available at: www.genome.gov/sequencingcosts. accessed July 2012)

Initial instrument costs for NGS platforms range from around £70,000 for benchtop models to over £500,000 for the larger systems.

Calculating the cost for sequencing a genome requires knowledge of the size of that genome and the required 'sequence coverage' (i.e., 'sequence redundancy') to generate a high-quality assembly of the genome given the specific sequencing platform being used. The costs quoted in Figure 11 will vary according to platform, sample throughput and level of automation within individual laboratories and have been calculated by the National Human Genome Research Institute (NHGRI) to include:

- Labour, administration, management, utilities, reagents, and consumables
- Sequencing instruments and other large equipment (amortized over three years)
- Informatics activities directly related to sequence production (e.g., laboratory information management systems and initial data processing)
- Library construction (required for preparing DNA to be sequenced)
- Submission of data to a public database

Multiple samples can be multiplexed in to a single run on the majority of platforms, which reduces the cost per sample at the expense of depth of sequencing per sample. Intense demand, continued innovation and more widespread use from both industry and academia will lead to further lowering of costs in the 2nd generation methods and further development of the 3rd and future generation methods.

The main steps in the 2nd generation sequencing methods are DNA extraction, construction of a DNA library, library amplification and sequencing, data analysis and validation of the results. Because each sequencing run can generate up to 600 gigabases (Gb) of data the demand for data analysis and storage is high.

Several 2nd generation sequencing approaches are available which all depend on unique sequencing chemistries. These in turn necessitate different methods for library preparation and analysis of sequencing data. Numbers of reads and length of individual reads also varies greatly between platforms. These differences present significant challenges when attempting to compare platform performance and data quality, and to develop appropriate standardisation approaches.

Examples of market providers include Roche, Illumina and Life Technologies and the different sequencing chemistries of selected platforms are summarised in table 1. Specifications are given for the purpose of comparison only; this is a rapidly evolving field with advances in output in terms of number of reads and read length constantly being reported. Short-read platforms (e.g. Illumina, SOLiD) typically produce a greater number of reads and approaches offering longer read length (e.g. Roche) combined with higher accuracy generally provide a means of improved species identification.

Platform	Clonal amplification method	NGS Chemistry	No. of reads/run	Read Length (bases)	Run time (days)
Roche/454	emPCR	Pyrosequencing	1 million	700	23 hours

GS FLX Titanium XL+				(average) 1000 (max)	
Illumina Genome Analyzer IIx	Solid-phase bridge amplification	Sequencing by reversible termination	320 million single-end/ 640 million paired-end	50 (average) 150 (max)	7 days (SE), 14 (PE) (151 cycles)
Illumina HiSeq 2000			3 billion single-end/ 6 billion paired-end	50 (average) 100 (max)	8.5 days (2 x 100bp)
Life Technologies SOLiD 4 (5500)	emPCR	Cleavable probe sequencing by ligation	100 million	50 x 35 (PE) (average) 75 x 35 (PE) (max)	3.5 days

TABLE 2 SPECIFICATION OF MAJOR NGS PLATFORMS (adapted from (DEVONSHIRE AS 2012)

NGS methods require template at relatively high concentrations compared to PCR methods. This is an issue when dealing with DNA extracted from food samples which may have low concentrations of organisms present and high fat, starch or protein levels that can have a direct affect on the yield. Consequently a culture/enrichment step is often required to generate sufficient material.

Sequencing strategies vary according to the initial question being asked and typically consist of sequencing entire genomes (or transcriptomes) from pure cultures or performing PCR amplification of targeted regions (such as rRNA regions) followed by sequencing of the amplicons. More recent developments enabled by the high throughput nature of the platforms include direct sequencing of clinical or food samples containing mixed and complex microbial communities. Considerations such as coverage or depth of sequencing required (number of individual reads of a certain region) and accuracy of sequencing will inform platform selection and the sequencing strategy to be followed.

The following sections will describe some of the major platforms in more detail.

7.3.2.1. Roche Genome Sequencer

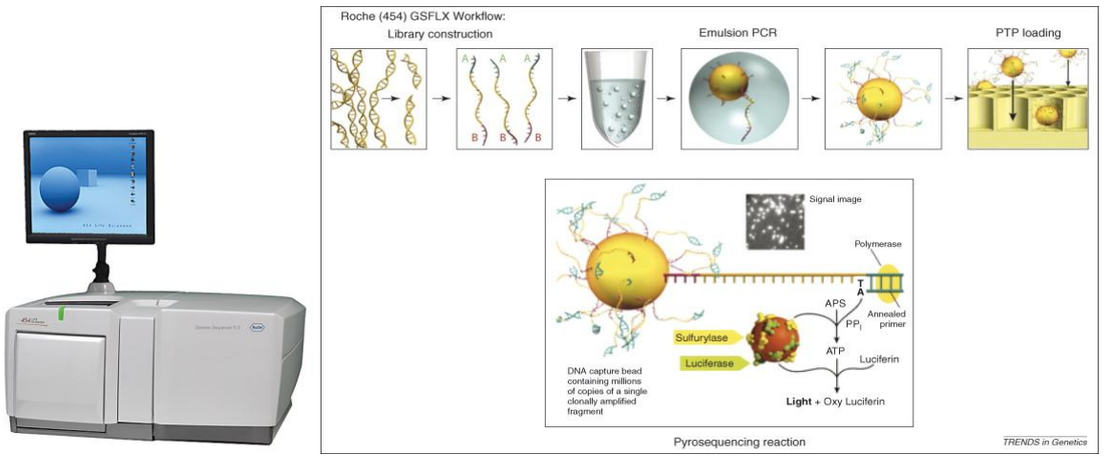


FIGURE 12 OVERVIEW OF THE ROCHE GS SYSTEM (obtained from http://www.roche.co.uk/)

The Roche GS systems (also widely known as 454 sequencing) use DNA polymerase to synthesise a complementary DNA strand and work on the principle of pyrosequencing, where the release of pyrophosphate upon nucleotide incorporation results in luminescent signal output.

Libraries consisting of fragmented target DNA are amplified en masse on the surfaces of hundreds of thousands of droplet encapsulated agarose beads using emulsion PCR (emPCR). Each bead captures a single library fragment which are then separated in to micelles (water and oil) containing PCR reagents. These are then applied to the surface of the picotiter plate (PTP) which consists of single wells in the tips of fused fibre optic strands that can each hold a single agarose bead. Imaging of the PTP following cyclical addition of each of the four base nucleotides serves to measure light emission as a consequence of nucleotide incorporation (Petrosino, Highlander et al. 2009).

The GS system typically generates longer reads than the Illumina and SOLID platforms, which is advantageous for taxonomic studies.

Accurate quantification of homopolymeric sequences may be problematic for GS sequencing as the linearity of response can exceed the level of detector sensitivity, a recognised issue with pyrosequencing, leading to insertion/deletion (indel) errors

7.3.2.2. Ilumina

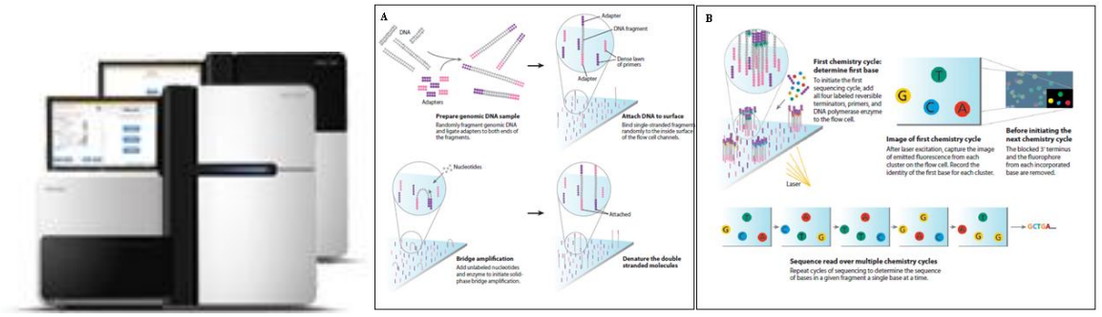


FIGURE 13 OVERVIEW OF THE ILLUMINA TECHNOLOGY (obtained from http://www.illumina.com/)

The Illumina chemistry is also based on sequencing by synthesis. However, unlike the 454 and SOLiD technologies which employ emPCR, the Illumina NGS platforms achieve target amplification in a flow cell by “bridge” amplification which relies on captured DNA strands “arching” over and hybridising to adjacent oligonucleotide anchors. Multiple amplification rounds convert single-molecule DNA template to clonally amplified arching clusters, with each cluster containing in the region of 1000 clonally amplified molecules.

Illumina sequencing works on the principle of reversible termination with each sequencing cycle involving the addition of DNA polymerase and a mixture of four differently coloured reversible dye terminators followed by imaging of the flow cell. The terminators are then unblocked and the reporter dyes cleaved and washed away. Following sequencing from a single end of the template, paired-end sequencing can be achieved by sequencing from an alternate primer on the reverse strand of the template molecule.

The Illumina NGS technology typically generates shorter read lengths (up to 100bp) but has one of the highest data outputs (numbers of reads) of the major platforms. Base-call accuracy decreases with increasing read length on Illumina NGS platforms because of “dephasing noise” due to under- or over-incorporation of nucleotides or failed terminator removal with successive cycles leading to the generation of a heterogeneous target-strand population within the cluster. This heterogeneity decreases signal purity and reduces precision in base calling, particularly towards the 3’ ends of a read (Dohm, Lottaz et al. 2008; Dohm, Lottaz et al. 2008).

7.3.2.3. Life Technologies SOLiD

FIGURE 14 OVERVIEW OF SOLID TECHNOLOGY (obtained from http://www.lifetechnologies.com/uk/)

The SOLiD (Sequencing by Oligo Ligation and Detection) platform employs a sequencing process catalyzed by DNA ligase. Similar to the Roche 454 technology, DNA fragments are amplified by emPCR while bound to the beads, after which the beads are covalently bound to the surface of a specially treated slide which is then placed into the fluidics cell of the instrument. Sequencing is initiated by the annealing of a universal sequencing primer to the adapters of the fragment library followed by addition of semi-degenerate fluorescently-tagged 8mer-oligonucleotides, which are ligated to the universal primer by DNA ligase when complementary to the sequence of interest. Following imaging of the slide, a subsequent cleavage step removes the sixth through to the eighth base, plus the fluorescent tag of the ligated 8-mer and a further nine ligation rounds performed. The sequencing strand is then denatured and washed away and a second round of sequencing is performed using a universal primer of (n-1) bp in length. A further three rounds are performed so that each base of the interrogated fragment is sequenced twice.

The SOLiD technology has a high accuracy rate for raw reads (>99.9%) due to the double interrogation of each base, and that procedure requires a lower volume of oversampling in order to reach a threshold value of confidence for base calling.

7.3.3. Next generation sequencing – Third generation

A third generation of sequencing platforms which offer further advantages in terms of reductions in sequencing time and cost are emerging. Examples include “bench-top” versions of the larger second generation instruments, single molecule sequencing platforms and non-optical sequencing technologies.

Bench top instruments with chemistries similar to their larger counterparts described in the previous section, such as the Roche GS Junior, Illumina MiSeq and Life Technologies Ion PGM and Ion Proton Sequencer, can provide a way to perform NGS which is cost-effective for diagnostic laboratories. Several platforms can generate results in under a day. Benchtop sequencers also offer more streamlined workflows for data analysis and generate more manageable quantities of data, which may be more suitable to the clinical laboratory end-user for future diagnostic applications.

The second generation sequencing instruments relied on PCR to create sufficient mass of the DNA fragments to be analyzed. The PCR step adds extra expense and time to the process and the amplification process can introduce bias and errors into the amplified products. These errors are perpetuated in the DNA sequence obtained, which ultimately increases the error rates of these technologies. Development of third generation sequencing platforms that can analyze individual single molecules of native DNA without the need for PCR amplification overcomes many of these problems. Longer read lengths are also typically obtained with the single molecule sequencing platforms.

7.3.3.1. Helicos

The first single molecule sequencing machine was developed by Helicos. The Helicos system performs “true single molecule sequencing” (tSMS) of DNA molecules captured on its flow-cell surface. In principal the Helicos approach is similar to that of Illumina, where reversible cy-5 labelled terminators for each of the four nucleotides are cyclically presented for incorporation into the extending DNA strand, a complete cycle of four nucleotides is termed a "quad". Typically 20 to 30 quads are performed resulting in read-lengths of 25-55 bases (average 35). The key differences with Illumina being that there is no target amplification and single molecule fluorescence is detected. The single molecule approach eliminates the dephasing problem inherent in the Illumina platform but the small signal from single molecule fluorescence leads to an increased per-nucleotide error rate and a much higher frequency of missing nucleotide calls that manifest as single nucleotide deletions in the resulting sequence.

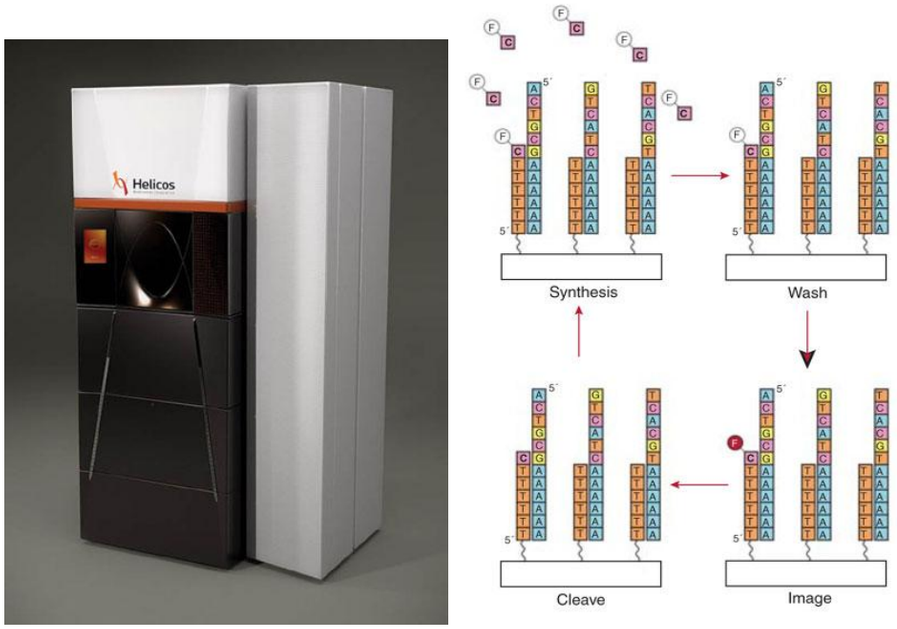


FIGURE 15 OVERVIEW OF THE HELICOS SYSTEM (obtained from http://www.helicosbio.com/)

7.3.3.2. PacBio RS system

The PacBio RS system uses “single molecule real-time detection” (SMRT) which detects the fluorescence of a labelled nucleotide as it is incorporated into the growing DNA strand. Single DNA polymerase molecules are tethered at the bottom of an optical chamber known as a zero mode waveguide (ZMW) with 75,000 individual perforations. As the fluorescent label is initially attached to the dNTP phosphate group it is cleaved during nucleotide incorporation, there is no need for reverse terminators and as each nucleotide is separately labelled there is no need to cyclically alternate the availability of nucleotides. Thus, this instrument can obtain sequence information at the processivity rate of DNA polymerase, which is several hundred bases per second. The PacBio RS is designed to produce average read lengths greater than 1,000 base pairs.

FIGURE 16 OVERVIEW OF THE PACBIO SYSTEM (obtained from http://www.pacificbiosciences.com/)

7.3.3.3. Oxford Nanopore Technologies

Oxford Nanopore Technologies have developed a single molecule sequencing platform based upon nanopore sensing. An ionic current is passed through a nanopore by setting a voltage across this membrane. If an analyte passes through the pore or near its aperture, this event creates a characteristic disruption in current. By measuring that current it is possible to identify the base in question as it passes through the pore. Single-use, self-contained cartridges that include all the reagents required to run an experiment simplify the workflow process. The process has also been miniaturised into a portable device for electronic single molecule sensing that plugs directly into a laptop or desktop computer through a USB port (MinION). A single MinION is expected to retail at less than £600.

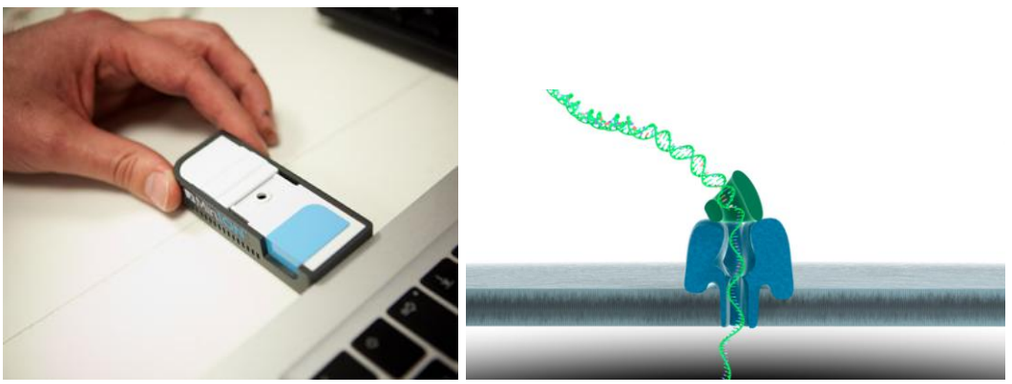


FIGURE 17 OVERVIEW OF THE OXFORD NANOPORE SYSTEM (obtained from http://nanoporetech.com/)

7.3.3.4. Ion Torrent and Ion Proton (Life Technologies)

The Ion Torrent (Life Technologies) platforms constitute a shift in technology from optical-based sequencing systems, measuring fluorescence or luminescence output, to monitoring release of hydrogen ions during DNA synthesis in a semiconductor-sensing device. The chemistry is similar to the 454 technology except that hydrogen ions produced when a nucleotide is incorporated into the growing DNA strand by DNA polymerase are detected on this platform (Perkel 2011). Micro-wells are incorporated on to an ion-sensitive layer on an ion sensor. An ion-sensitive field-effect transistor (ISFET) is used as the detector to measure the released H+ concentration (Rothberg, Hinz et al. 2011).

The Personal Genome Machine (PGM) sequencer can generate up to 8 million reads and sequencing runs are performed in less than one hour with read lengths anticipated to exceed 400 bp by end of 2012.

The recently launched Ion Proton platform brings costs and turnaround times down further.

FIGURE 18 THE ION TORRENT AND ION PROTON PLATFORMS (obtained from http://www.lifetechnologies.com/uk/)

7.4. NGS Data analysis and informatics requirements

Each NGS sequencing run can generate over 600 gigabases (Gb) of data, therefore the demand for data analysis and storage is high. For all but the most simple experiments expert bioinformatics input is required.

Initial steps in the analysis procedure include data quality control (QC) and filtering to ensure that the data passes a variety of quality thresholds. These may include the total numbers of reads generated and the average read length, and if bi-directional sequencing is performed the numbers of reads in each direction should be approximately equal.

For microbial sequencing, the proportion of the reads that can be assigned to a particular taxonomic level (e.g. genus, family, class, phylum, etc) is an important QC check.

Data are also typically filtered and cleaned to remove artefacts by removing reads with low individual quality scores, removing low complexity sequences, allowing no more than 2 mismatches between sequence-read and PCR primer and removing sequences with Ns and/or homopolymers of longer than 8.

For 16S amplicon sequencing, alignment against 16S databases and elimination of sequences that align over less than 200 bp of their length or fail to align to the target 16S region can be removed. Chimeric sequences can also be removed using software such as ChimeraSlayer (Haas, Gevers et al. 2011).

Following data QC and filtering a key objective in most NGS analyses is to assign equivalence to sequence reads to identify sample source and taxonomic grouping. This either entails aligning reads to a common comparator such as is contained in a reference database of sequences. The need to accommodate the high rate of sequence errors in NGS data and also genuine sequence differences such as polymorphisms add substantially to the complexity and thus computational time of the task of sequence alignment and identity determination.

For metagenomic analysis operational taxonomy units (OTUs) are typically defined first by clustering reads according to sequence similarity. A similarity (or distance) threshold is chosen to determine level of grouping.

A variety of databases is then available to assign OTUs to taxonomic groupings. This is done by looking for sequence similarity between a representative sequence from each OTU versus the reference databases. rRNA gene sequences (e.g 16S) are widely used for phylogenetic studies and there are several tools for analysing and comparing 16S data, such as BLAST, SILVA (Pruesse, Quast *et al*. 2007), RDP (Cole, Wang *et al*. 2009), GreenGenes, QIIME, MEGAN, GOLD and MG-RAST.

Appendix 3 describes some of the commonly used publicly available databases and search tools.

7.5. Uptake of NGS technologies and current expertise

The first full microbial genome (*Haemophilus influenza*) was sequenced in 1995 using Standard Sanger sequencing (Fleischmann RD 1995) and since then, with the advent of NGS technologies the numbers of microbial genomes sequenced has increased at an exponential rate with over 4,000 individual microbes sequenced to date.

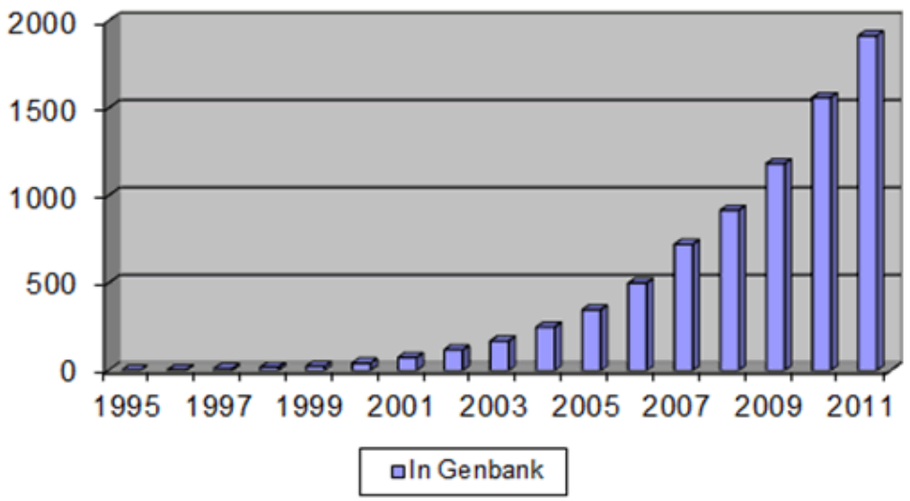


FIGURE 19 NUMBERS OF COMPLETELY SEQUENCED MICROBIAL GENOMES (UP TO OCTOBER 2011) (obtained from http://www.genomesonline.org/)

As sequencing instruments become ever faster and cheaper they are also starting to spread beyond big centres into reference laboratories, clinics and smaller laboratories. The following graph shows the dramatic rise in publications (from PubMed) based on using the search terms “next generation sequencing” or “high throughput sequencing” since the year 2000.

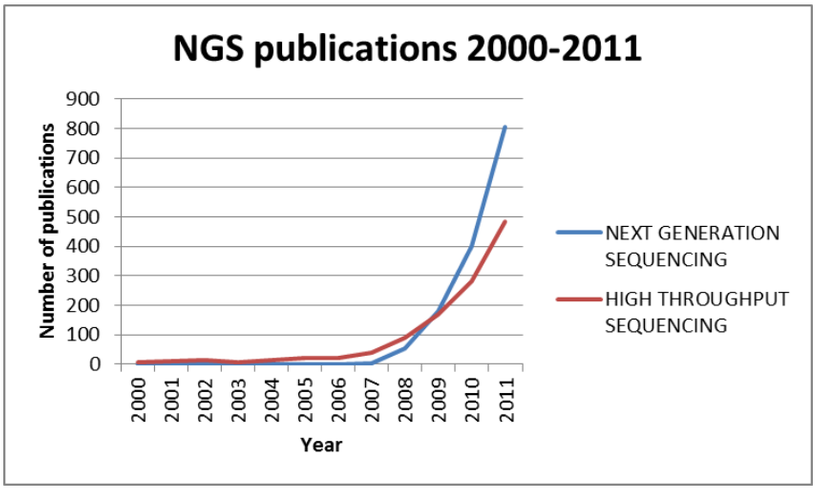


FIGURE 20 NUMBERS OF NGS PUBLICATIONS PER YEAR

The number of NGS instruments in the World is mapped on omicsmaps.com by James Hadfield (Cancer Research UK, Cambridge) and Nick Loman (University of Birmingham). According to this resource as of May 2012 there were 142 high throughput sequencing instruments in the UK, which breaks down into the various platforms; 42 Illumina HighSeq, 40 Illumina Genome Analyser IIx (GA2), 25 Roche 454, 16 ABI SOLiD, 9 Ion Torrent, 8 Illumina MiSeq and 2 Pacific Biosciences. Instruments are primarily housed in centralised sequencing facilities, such as the Wellcome Trust Sanger Centre and The Genome Analysis Centre (TGAC), as well as in individual Higher Education Institutes (HEIs) such as at the University of Birmingham, Bristol, Exeter, Glasgow, Manchester, Liverpool and Warwick. The establishment of sequencing facilities are funded by funding councils, Biotechnology and Biological Sciences Research Council (BBSRC), Natural Environment Research Council (NERC) and Medical Research Council (MRC), charities, universities and a Regional Development Agency.

FIGURE 21 UK LOCATIONS OF NGS INSTRUMENTS (obtained from http://omicsmaps.com/)

Appendix 4 lists the major facilities in the UK with NGS expertise and also lists experts in the area of food microbiology testing, epidemiology and clinical microbiology.

7.6. Reported application of NGS in outbreak settings

The majority of studies to date using NGS methods in relation to foodborne outbreak investigation have required culture to obtain pure isolates for sequencing. These studies (described below) have reported the potential of such approaches for determining virulence factors, pathogenicity and strain type for epidemiological investigations to trace outbreak source and spread, and drug resistance for clinical decision making. The ability of these methods to determine the relationship between isolates which may be otherwise indistinguishable underlines their great potential in epidemiological investigations of outbreaks of foodborne pathogens.

The outbreak of *E. coli* O104:H4 in 2011 demonstrated the potential of sequencing technologies for use in outbreak investigations. This outbreak occurred at a point in time when many groups around the world had the capability to perform NGS and consequently isolates from the outbreak were sequenced by many groups on multiple NGS platforms in near real-time as the outbreak was occurring. Not only were NGS platforms widely used in near real-time for the first time, but the method for the outbreak investigation was novel in that groups performing the sequencing shared data and jointly analysed the datasets thanks to the parallel developments in cloud computing and social media. In a review of the investigation process published shortly after the outbreak entitled "*E.coli* O104:H4 Social media and the characterization of an emerging pathogen" Casey *et al* (Casey PG 2011) state:

"*What has differentiated the investigation of this outbreak from those of previous outbreaks of foodborne illness is the method of investigation. Rather than a meticulous process involving the acquisition of data by a single official body followed by peer review and subsequent publication in scientific journals, the characterization of the source of the German outbreak has proceeded at an astonishing pace, with assorted tweeters and bloggers across the globe collaborating online in a near–real time effort to rapidly understand the genetic makeup which made the outbreak strain so virulent. Crowdsourcing such as this has been defined as "a distributed problem solving model" and this global bioinformatic collaboration has produced answers faster than many would previously have thought possible."*

Platforms used in the investigation included the PacBio, Illumina, Roche 454 and Ion torrent systems. (Rasko, Webster et al. 2011). (Mellmann, Harmsen et al. 2011) (Rohde, Qin et al. 2011). Researchers at BGI in Shenzhen, China, obtained samples of the outbreak strain from the University Medical Center Hamburg-Eppendorf. Within five days, BGI had completed seven runs on an Ion Torrent instrument and released the sequencing reads into the public domain. This triggered crowdsourced analyses by bioinformaticians across the globe. A day later, a de novo assembly of the genome had been produced by a bioinformatician in the UK (http://www.bbsrc.ac.uk/news/research-technologies/2011/111108-f-wisdom-of-crowd.aspx) and within a week, over 20 entries had been filed on a new website dedicated to genomics of the strain, revealing details of its pathogenic potential and evolutionary origins. BGI subsequently used an Illumina HiSeq to produce an improved assembly.

In parallel, the University of Münster and Life Technologies also sequenced the genome of an isolate from the outbreak using Ion Torrent sequencing. Two other research centres sequenced isolates from the outbreak using 454 technology—the Göttingen Genomics Laboratory and UK's Health Protection Agency (HPA). Finally, Pacific Biosciences used their platform to sequence not only the outbreak strain, but also 11 other related strains.

An editorial in Nature Biotechnology (Annonymous 2011) analysed the contribution of sequencing to the outbreak investigation. They concluded that in this case, the sequencing made little or no difference to the actual management of patients or of the outbreak as a whole. However, the ability to obtain genome sequences from outbreak strains within days of isolation allowed the pathogen to be characterized at extraordinary speed. Similarly, an open attitude to data release meant that researchers and public health scientists could immediately exploit, and even add to, sequence-based information, using the existing well-defined collection of HUS-associated type strains.

The commentary also noted that the outbreak provided an opportunity to evaluate the range of competing sequencing technologies and associated bioinformatics pipelines, with this strain being sequenced on more platforms than any other organism. Despite this, a finished genome for the strain is still lacking and there is no simple path from genome sequence to an understanding of virulence or transmissibility. For example, in this outbreak, the genome sequence revealed a list of potential adhesins, but proving their roles, if any, in adhesion to foodstuffs or to the human gut will require months or years of detailed investigation in the laboratory.

The editorial concluded by discussing the potential of sequencing to track the evolution of strains in outbreaks and highlighted the fact that high-quality sequence is needed to detect true variants for genomic epidemiology. NGS approaches currently introduce too many errors into draft sequences—at rates as high as 1 in 100,000 bases—to make real variation distinguishable from base-call errors, at least without confirmation by Sanger sequencing. Improvements in the sequencing accuracy of the technology in future generations of instruments may help to overcome this problem.

Examples of the use of sequencing in previous outbreaks have also been reported. The causative agent in a major outbreak of salmonellosis in the US in 2009-10 could not be distinguished from other strains of *Salmonella* enterica using PFGE. However, NGS analysis identified single nucleotide polymorphisms and acquisition of a phage with homology to several coding segments of an Enterobacteria phage D6 (Lienau, Strain et al. 2011). The NGS data allowed the source of the outbreak to be identified and a single food facility to be identified as the origin.

Investigation into the origin of an outbreak strain of *Vibrio cholerae* in Haiti was performed using the PACBIO RS third generation NGS instrument (Chin, Sorenson et al. 2011) Genomic analysis revealed that the strain associated with the Haitian outbreak was likely to have been introduced from a distant geographical location by human activity as it was closely related to a strain isolated in Bangladesh, variant *V. cholerae* EI Tor O1, and only distantly related to current isolates circulating Latin America and East Africa. Therefore the authors concluded that the cholera outbreak was not likely to have resulted from climatic events prior to the outbreak. Further,

whole genome analysis of this outbreak has identified single nucleotide polymorphisms (SNPs) to inform the epidemiological spread of these EI Tor O1 isolates which was originally thought to have been quite diverse. In addition to identifying SNPs, the acquisition of genetic elements involved in drug resistance was mapped, identifying when this specific family of genetic elements was acquired and how it influenced the spread of the pandemic. This clearly demonstrates the contribution of NGS methods into the epidemiological study of outbreaks.

Current studies highlight the diagnostic and epidemiological potential of NGS methods in foodborne disease. NGS technologies can be used in strain typing methods to develop a high resolution and high throughput method. This removes the need to sequence whole genomes and instead allows a more directed approach by sequencing defined genomic regions. These target gene sequences are amplified by PCR prior to sequencing. The application of this technique has been demonstrated in Salmonella strains as an alternative to conventional multilocus sequence typing (MLST) method (Singh, Foley et al. 2012).

More recently methods using metagenomic approaches have looked to remove the culture step by performing analysis of entire microbial communities directly from uncultured samples. This complementary approach can provide a more realistic representation of pathogens in the environment, many of which cannot be cultured and allows identification of co-infections and microbial community interactions. However, this benefit has to be balanced against the additional complexity of sequencing potentially hundreds or thousands of different organisms simultaneously, and also the generation of data from both viable and non-viable organisms. Metagenomic sequencing strategies include sequencing of 16S rRNA PCR amplicons produced directly from clinical or food samples, or entire “shotgun” sequencing of total DNA extracted from clinical or food samples. As microbes function in complex communities with many interdependencies the additional information gained through the identification of co-infections and other interactions should help to inform better interpretation of clinical manifestations and epidemiological investigations.

The Human Microbiome Project (HMP) aims to characterise the ecology of human-associated microbial communities. A recent ground-breaking publication by the HMP (Consortium 2012) analysed a large cohort and set of distinct, clinically relevant body habitats using metagenomic sequencing directly from clinical samples. The results highlighted the range of structural and functional configurations normal in the microbial communities of a healthy population. This baseline information should enable future characterisation of the epidemiology and ecology of the human microbiome in healthy and disease states. Figure 22 illustrates the diversity of the human microbiome.

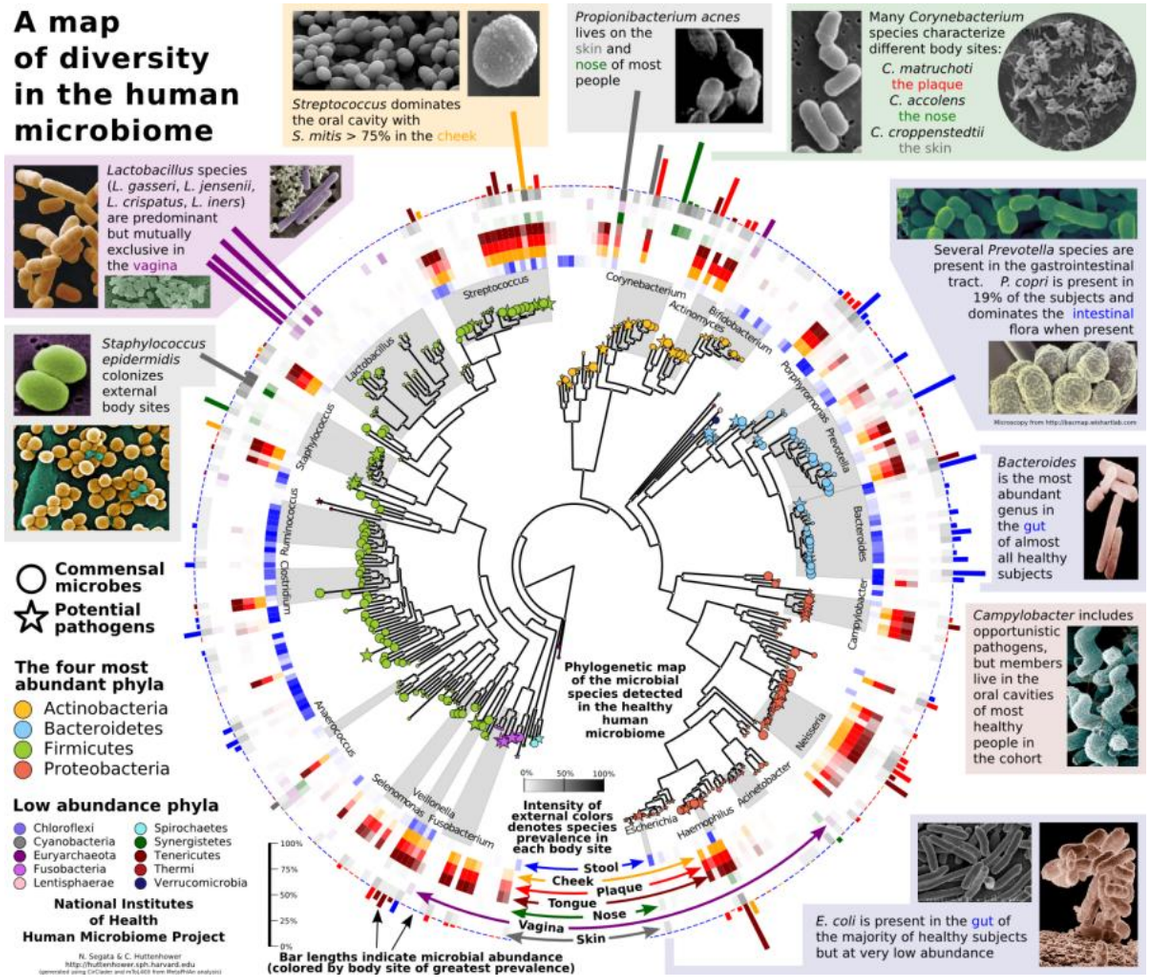


FIGURE 22 DIVERSITY OF THE HUMAN MICROBIOME DETERMINED BY METAGENOMIC SEQUENCING (taken from http://huttenhower.sph.harvard.edu/metaphlan/)

8. Potential role of emerging technologies in outbreak management

The following sections outline where current disease outbreak management might be enhanced by incorporating molecular testing earlier in the process (either alongside or in place of culture approaches). This will require formal evaluations to be performed. If insufficient evidence exists studies should be undertaken to generate data comparing molecular and culture-based approaches.

8.1. Establishing that there really is an outbreak

Outbreaks are often detected through routine surveillance of laboratory-confirmed cases of infection. In the UK microbiological diagnosis of cases of gastrointestinal infection still tends to rely on traditional culture. Isolates of certain organisms, e.g. salmonellas or Vero cytotoxin-producing *Escherichia coli*, will be forwarded to a reference laboratory for confirmation of identity, phenotyping and genotyping. These examinations tend to be conducted in a hierarchical manner. At times when the reference laboratories are very busy this can result in delays of several days or weeks before an outbreak is detected.

The challenge is to implement rapid detection and typing methods that are optimal for detecting, through routine surveillance, clusters of isolates (cases) that may be associated with either a point source exposure or a continuous source in as close to real-time as possible.

The implementation of rapid, high throughput testing as facilitated by molecular methods may allow samples to be analysed more quickly and in a more cost effective manner potentially speeding up routine surveillance of isolates and allowing faster recognition of a potential outbreak.

8.2. Confirming the diagnosis

This step involves collecting clinical samples from people with illness and processing them in a clinical laboratory. Currently, however, there is variation in clinical laboratory practice across the NHS. For example “No Pathogen Detected” in one laboratory might mean that *Salmonella*, *Campylobacter* and *Shigella* infections have been ruled out using traditional methods whilst in another *E. coli* O157, *Cryptosporidium* and *Giardia* infections have also been ruled out. Furthermore, “No Pathogen Detected” in either scenario is not the same as “No Pathogen Detectable.” Certain pathogens are sought only in specific circumstances and these can vary between laboratories too – the classic example being the variable approach to norovirus diagnosis. Furthermore, different methods/media are available for clinical use. Using molecular methods for primary diagnosis should promote more standardisation than exists at present. Implementation of emerging molecular approaches capable of screening for multiple pathogens simultaneously may also improve diagnostic accuracy.

A further consideration is the low diagnostic yield from traditional microbiology. Around 90% of all stool samples submitted to clinical laboratories and examined using traditional methods will produce a negative result. Yet the second study of infectious intestinal disease in the community (IID2 Study) demonstrated that it is possible to increase diagnostic yield substantially using molecular methods (Tam *et*

al, 2012). Subject to affordability, it should also be possible to incorporate a pathogen discovery step into routine surveillance using whole genome sequencing.

Finally, whatever the methods employed, they need to be able to differentiate outbreak cases from background cases reliably and rapidly (hours or days rather than days to weeks). Molecular approaches are likely to be the only realistic strategy for fulfilling this requirement.

8.3. Creating a case definition

The case definition usually comprises simple and objective clinical criteria with limitations on time, place, and person. Sometimes different levels of case definition are needed e.g. probable cases (patients with similar symptoms) and confirmed cases (where a laboratory diagnosis is added to the definition for a probable case). However, with notable exceptions like the US where molecular typing is employed routinely in PulseNet, there are relatively few instances when highly discriminatory results are available in real-time as opposed to being applied post hoc. The greater the level of discrimination achieved the more precise is the case definition but this needs to be achieved in as close to real-time as possible. The more precise and discriminatory the case definition, the greater the chances of detecting a statistically significant association between illness and consuming a contaminated food. Several of the rapid and highly discriminatory approaches discussed in this review have the potential to provide information in near real-time.

8.4. Generating and testing hypotheses for exposure

This involves collating information about symptoms, circumstances, and diagnosis to form hypotheses about the cause of the outbreak, which can then be tested using analytical epidemiological methods like case-control, case-case or cohort studies. However, as can be seen from comments above, investigators may need to question cases about exposure to contaminated food that might have happened six to eight weeks ago, increasing the potential for recall bias. If employing molecular methods can reduce the time from outbreak recognition to this type of investigation this will have the potential to reduce recall bias. The greater the precision in the case definition the greater are the chances of detecting a real statistical association between exposure and disease.

8.5. Considering what additional evidence is needed

Alongside the epidemiological investigations described above, Environmental Health Officers will assess food preparation practices, inspect premises and obtain samples of food for testing, if available. They may also swab food preparation surfaces. Finding organisms with indistinguishable molecular profiles in clinical samples and in food or environmental samples provides robust (and in some instances actionable) evidence of a link between clinical cases and an implicated food/food premises.
Many of the currently available traditional and PCR based approaches are able to discriminate down to the species level, but are unable to reliably discriminate down to the strain or sub-type level. The newer approaches discussed in this review based on sequencing will be capable of providing this information.

8.6. Source attribution

As well as improved outbreak detection, molecular markers are also useful for source attribution. This involves characterization of isolates of a specific pathogen by phenotypic and/or genotypic subtyping methods (Pires, Evers et al. 2009). The aim is to compare the subtypes of isolates from different sources (e.g., animals, food) with those isolated from humans. It is facilitated by strong associations between some of the dominant subtypes and a specific reservoir or source, providing a heterogeneous distribution of subtypes among the sources. Subtypes isolated exclusively or almost exclusively from one source are considered to be ‘‘indicator subtypes,’’ and the human infections caused by each indicator subtype are assigned (attributed) to that specific source. The relationship between the relative occurrence of each indicator subtype in the source and the incidence of human infections caused by that indicator subtype is then established. Finally, human infections caused by subtypes found in several sources are assigned to specific sources proportional to the occurrence of the indicator subtypes. Hence methods that are useful for outbreak investigations should be capable of being used for associated analyses like source attribution and source tracking.
As discussed above, newer molecular approaches based on sequencing will be able to provide detailed information on subtypes and sequence similarities.

9. Barriers to uptake of molecular approaches as front-line tools

Considering time to results alone (hours to days for molecular approaches as opposed to days to weeks for traditional microbiological methods), molecular methods have clear advantages. The additional ability of some of the newer approaches to reliably identify down to the strain or serotype level and provide information not previously achievable using standard approaches offers further obvious advantages. However, molecular approaches have not yet been universally adopted as a front line tool in outbreak situations. Barriers to their uptake include:

9.1. Cost

Generally, each molecular system outlined in previous sections will require an initial investment in instruments and computing system. For some of the higher throughput instruments these items may pose a significant cost to the purchaser. In addition to the one off capital costs are the continuous consumable costs, which can be prohibitive for smaller laboratories wishing to adopt the technologies. The question of being able to use the same equipment for different microorganisms also becomes an issue, as some specialised systems require the use of assay-specific commercial kits and may not therefore be able to detect a wide enough suite of pathogens for one laboratory, potentially requiring investment in two or more systems. By contrast, current culture-based systems rely on simpler technology, i.e. keeping different culture media at a specific temperature is achievable with any incubator.

A recent survey of food laboratories carried out by Applied Biosystems in 2009 allowed an insight into perceptions of molecular based technologies as well as the main factors that influence the decision to introduce new technologies. 80% of the laboratories participating in the survey stated that initial investment was the key barrier to the introduction of alternative methods within their organisation. In addition to the capital outlay required to purchase the instrument, 65% of the respondents

stated that the price of the consumables was another factor preventing the introduction of other technologies into the laboratory (personal communication, Applied Biosystems).

9.2. Re-training and laboratory infrastructure

The current situation is that food analysis laboratories have expertise and accreditation in culture-based microbiology. The task of re-training staff to use molecular methods, validating the methods and gaining accreditation is not a small one, and the cost of this may be prohibitive. In addition, for the newer NGS approaches additional skill sets are required for data analysis and bioinformatics. These skills are not easily transferrable and recruitment of specific expertise will often be required unless the service is outsourced to specialist laboratories. A sensible laboratory manager will also run any new system alongside the old system for a period of time to ensure that the new system is able to perform as well as the old system. Many laboratories will not have the spare capacity to be able to run two sets of analysis concurrently. Aside from the need to re-train staff in molecular biology techniques, the laboratory itself may need to be upgraded to be able to handle DNA confidently with an acceptable risk of contamination.

9.3. Ensuring “fitness for purpose” and comparability

As discussed earlier in this review (section 5) it is challenging for laboratories to implement molecular approaches such as PCR when appropriate guidance and mechanisms for validating assays and assessing comparability and “fitness for purpose” are lacking. This challenge will be magnified when some of the emerging data-rich sequencing approaches are considered. The following sections will discuss some of these issues further.

9.4. Method validation and laboratory accreditation

A standardised approach to the comparison and validation of new methods is available (ISO 16140) and this can be used, and indeed is required to be used in European Commission Regulations, before food testing using ‘Alternative Methods’ can be used. These mechanisms can be certified by third-party ‘Validation Bodies’ such as AOAC, MicroVal or AFNOR. The idea behind such validations is to perform a rigorous third-party evaluation of the technique in question to achieve a level of accuracy comparable to or better than current standard tests across a range of food matrices. This in itself is challenging when molecular approaches are often measuring different parameters to standard culture approaches, making comparisons difficult. However, the resulting certificate of validation allows one to have some level of confidence in the method when applied in a real-life situation. This ability to gain third party accreditation has allowed the food industry to use molecular tests with some confidence. The number of kits listed in Appendix 2 shows that there is a market for molecular methods in the food industry even if uptake of these methods is by no means the norm.

The situation for clinical testing is somewhat different. Should a clinical laboratory decide to use a particular molecular method, it must be validated according to the procedures “Commercial and In-House Diagnostic Tests: Evaluations and

Validations" (HPA 2012) and "Guidance on the development and validation of rapid diagnostic tests that rely on nucleic acid amplification and detection (Saunders, Sharp *et al*. 2011). This procedure puts the onus of validating the method on the laboratory choosing to use it. Comparing the validation procedures for clinical diagnostics with those produced for the food industry is revealing. The HPA guidance has no specific requirement for any particular novel method to have been confirmed as accurate through an inter-laboratory study. The HPA guidance also gives laboratories the responsibility for choosing experimental methodology and the appropriate statistical analysis for any particular method. Compared with the food industry criteria, this approach allows much greater freedom to tailor experiments and analyses to suit a particular laboratory's workload. However, little guidance on choosing appropriate validation approaches is available and the resulting validation may not then be applicable to another laboratory wishing to use the same clinical diagnostic technique. By contrast, the food industry validation process provides a level of confidence in a technique applicable to any laboratory that follows exactly the same methodology as used in the validation.

If the emphasis remains on an individual clinical laboratory to demonstrate that any novel diagnostic test is "fit for purpose" including absorbing the costs of such an investigation, it seems likely that this will militate against adoption of novel methodology (even if it is likely to be superior in terms of time to results) due to the increase in workload, cost and paperwork required to be able to use it. In addition, reference methods and materials that facilitate such assessments are lacking (see following sections), which further hamper the process. A review of current practice to reduce the burden on individual laboratories and harmonise approval of novel tests in line with industrial guidelines should be undertaken in order to facilitate adoption of molecular approaches.

9.5. Lack of controls for defining a positive result

Using molecular methods for diagnosis allows use of low-volume samples and increases the sensitivity of detection (Tam, O'Brien *et al*. 2012). However, without appropriate controls, identifying an appropriate cut-off value to define a positive PCR result is challenging. When considering validation studies on new methods, where they are compared to conventional culture based procedures, there are occasions when a sample gives a positive result with an amplification based procedure and a negative by the culture based method. In such situations it is increasingly difficult to determine if this is a false negative by the culture based method or a false positive by the amplification based method.

It has been shown previously that a threshold cycle (Ct) value (the point in a qPCR assay where fluorescent signal can be detected above background) of <30 is a good indicator of infectious intestinal disease (IID) genuinely caused by norovirus and rotavirus (Phillips, Lopman et al. 2009; Phillips, Lopman et al. 2009). However, different PCR instruments, interfering sample matrices and real-time chemistries will result in Ct variability that will need to be controlled and accounted for through the use of appropriate standards and reference materials, which are currently lacking. In the absence of similar data on Ct value cut-offs for other organisms, a more sensitive cut-off value of <40 for other pathogens tends to be used. This might lead to an over-estimate of the contribution of those pathogens to disease if IID cases with high Ct values (low pathogen loads) are not actually due to infection with those organisms.

Often laboratories are left having to do considerable conventional microbiology trying to isolate a target organism from the sub-sample tested with the amplification based technique. The issue here is that microbiologists still require a cultured isolate to feel confident that they have definitely detected the target organism. Part of the future of molecular methods has to be increasing confidence of a positive in the absence of a cultured isolate. Studies to define better a positive result and demonstrate equivalence (or superiority) to culture based approaches will be required. Many molecular based assays simply detect the genomic DNA from pathogens which may be present whether the organism is viable or not. More advanced molecular approaches to detect evidence of viability (e.g. gene expression) may help to define better positive results.

9.6. Lack of controls for diagnosing mixed infections

Assays based on detection of multiple virulence genes can generate ambiguous results if a mixed culture is present to start with (Amar, East et al. 2007; Havelaar, Brul et al. 2010). Without appropriate controls it is difficult to determine whether co-infection is coincidental, whether it is due to shared routes of infection, or whether certain pairs of pathogens are more often associated with disease (as opposed to asymptomatic infection) (Tam, O'Brien et al. 2012). Further studies examining larger numbers of samples will be required to help clarify these issues.

9.7. Absence of isolates

As discussed above, a limitation of currently available molecular techniques is the failure to discriminate between viable and inactivated organisms if an initial culture step is not performed (Havelaar, Brul et al. 2010). However, future practical solutions to this are on the horizon as transcriptional activity around bacterial cell survival/death appears to reveal molecular markers for cell viability (Kort, Keijser et al. 2008). Reagents such as propidium monoazide (PMA) can also be used which selectively bind to free DNA and cannot pass through intact membranes of live cells. Research has indicated that this strategy could be used to selectively amplify DNA from live *Campylobacter* cells to assist in quantifying the level of organisms present in chicken carcass rinse (Josefsen et al 2010).

Despite these new technologies, culturing organisms will remain important. In the rush to adopt new technology the public health requirement for cultured organisms for characterisation including epidemiological sub-typing should not be forgotten (Jones and Gerner-Smidt 2012). What might be required is a rapid screen for diagnostic purposes using molecular methods followed by culture of the target organism to enable the public health response. This is important to maintain consistency with historical data and render trend monitoring meaningful.

9.8. Lack of molecular reference standards for platform comparability

As discussed in section 8.2 the implementation of “alternative” approaches requires validation to demonstrate equivalence or better to “gold-standard” approaches. To fully ensure confidence in molecular methods there is also a need to demonstrate comparability between molecular approaches as in outbreak situations data often

needs to be shared between different laboratories and countries. Molecular approaches not only differ from traditional culture approaches in that they are measuring different parameters (discussed in section 5), but they can also differ significantly from each other in terms of the analytes being measured and the method of measuring those analytes. This is particularly pertinent for NGS methods. Section 6.6 discussed typical QC steps in NGS data processing, but different laboratories are likely to use their own preferred QC metrics and thresholds, which can make comparing data generated in different laboratories on even the same platform difficult.

In addition to this, as discussed in previous sections, the different NGS approaches currently available produce sequence reads of different lengths and the numbers of reads produced can vary considerably. This makes comparing data across different platforms challenging. A further confounding factor is that each sequencing chemistry and platform will have unique sequence biases (e.g. tolerance of AT rich sequences, homopolymers and specific sequence motifs) and all have different levels of sequencing accuracy (rate of mis-incorporation of incorrect base) (Dohm, Lottaz et al. 2008). Even with fairly low mis-incorporation rates the sheer volume of data produced means that a significant number of sequencing errors will be present in a sequence. This not only makes comparing data difficult but also limits the ability of the approaches to discriminate down to the strain level as resolution is frequently dependent of the detection of a small number of single nucleotide sequence differences. Therefore, error rates and biases need to be well characterised as systematic sequencing errors may occur at the same base position on a high percentage of reads masking true sequence identity.

A 2011 publication by Suzuki *et al* (Suzuki, Ono *et al.* 2011) used a strain of *E. coli* to compare three different NGS platforms (Roche GS FLX System (FLX), Illumina Genome Analyzer (GA), and Applied Biosystems SOLiD system (SOLiD)). The sequence reads obtained were aligned to the complete genome sequence of *E. coli* DH1, to evaluate the accuracy and sequence bias of the methods. The fraction of “junk” data, which could not be aligned to the reference genome, was largest in the SOLiD dataset, in which about half of reads could not be aligned. Among datasets after alignment to the reference, sequence accuracy was poorest in GA datasets, suggesting relatively low fidelity of the elongation reaction in the GA method.

To enable comparability to be demonstrated well-characterised reference materials will need to be available including reference DNA from specific strains as well as panels of clinically validated samples.

Determining the identity of organisms analysed by sequencing will only be as good as the quality of the reference sequences in the databases. To exploit fully the potential of NGS technologies global genomic databases will need to be established with sequences from clinical samples performed in different laboratories deposited in to it. A recent initiative called the ‘100K Genome Project’ aims to accelerate foodborne pathogen analysis through the creation of a new database of 100,000 foodborne pathogen genomes to help speed identification of bacteria responsible for illness outbreaks. (http://www.foodproductiondaily.com/Quality-Safety/Five-year-100K-Genome-Project-aims-to-speed-foodborne-pathogen-analysis)

In short, industry standards need to be developed which ensure that sample preparation methods are robust and universal, analytical methods are standardised and high quality data are stored in a consistent format. Community driven efforts such as the Genomics Standards Consortium (GSC) (http://gensc.org/) are starting to address some of these issues.

The international measurement community is also working to improve the comparability of molecular methods by providing traceability to higher order reference methods and materials. The task of the Bureau International des Poid et Mesures (BIPM) (http://www.bipm.org/en/home/) is to ensure world-wide uniformity of measurements and their traceability to the International System of Units (SI). It does this with the authority of the Convention of the Metre, a diplomatic treaty between fifty-six nations, and it operates through a series of Consultative Committees, whose members are the national metrology laboratories of the signatory States. The Consultative Committee for Amount of Substance – Metrology in Chemistry (CCQM) oversees biological measurements and has recently established a steering group for microbiology to support harmonisation and inform efforts to support global comparability through traceability in this area. Global comparisons of specific measurement methods for food safety are currently being organised in close co-operation with the food community and linked to food safety regulators, professional organisations, documentary standards developers, kit manufacturers, industrial and reference laboratory microbiologists, food microbiology proficiency testing (PT) scheme providers, biological resource providers and national measurement institutes.

9.9. Lack of data reporting standards

Currently, molecular and especially NGS measurements are often generated using inconsistent approaches and interoperability of data is emerging as a major issue.

It has been reported that when NGS was used to investigate a *V. cholerae* outbreak researchers in a laboratory in Arizona were unable to compare the Nepalese strains they had sequenced with a Haitian strain sequenced in Harvard because the two groups used different instruments and the raw data were not compatible (Kupferschmidt 2011). Harmonisation of data outputs to enable compatibility will be essential going forwards.

9.10. Meaningful output

Finally it will be very important to be able to make sense of the output quickly, which means developing clinically and public health relevant reporting.
The management of large amounts of data generated by the newer molecular approaches requires careful evaluation with respect to reporting the findings implicating food specimens as potential sources of an outbreak. As highlighted in previous sections there are currently multiple databases and data analysis tools available, which may differ significantly in quality and analytical approach.

Databases would also need to include other important information relative to the outbreak to allow the results to be integrated with epidemiological data. Plans for data storage including management, transfer and security of data would need to be implemented. Costings for data storage also need to be considered. However, costs

could be somewhat mitigated by implementing cloud computing and storage, provided that adequate levels of data security, commensurate with handling sensitive information (e.g. personal identifiers), can be assured.

Ultimately there is also a need to develop easily interpretable visualisation software, which can be used by laboratory personnel who lack specific bioinformatics expertise. These programs could be developed to report useful clinical information such as the organism present and its corresponding strain type, drug resistance profile and important virulence determinants

10. Discussion

Molecular approaches are evolving rapidly and have the potential to reduce timescales and improve management of foodborne outbreaks as discussed in this review. Other reviews such as the 2007 HTA assessment of rapid methods also concluded that nucleic acid based methods show more promise to deliver fast and accurate results for a reasonable cost, despite the apparent limitations of such methods (e.g. investment in people and kit; Inhibitors in samples; comparable DNA extraction between samples; re-organisation of laboratories; lack of an isolate for confirmation).

Similarly, a recent HPA paper highlighted the advantages of using emerging NGS methods and also some of the challenges, including investment in kit, and concluded that their best approach for using NGS methods was to form alliances with already established centralised sequencing facilities at external organisations (HPA 2011).

With the advent of “personalised” bench-top sequencers in the last year or so, sequencing instruments are becoming ever faster, smaller, and cheaper and are now spreading beyond big centralized facilities into hospitals and smaller laboratories. However, although many laboratories are now implementing or considering implementing molecular assays such approaches are still largely being used later for confirmatory purposes rather than as first line approaches. The full potential of molecular approaches to improve and speed up outbreak investigations will only be fully realised if the approaches are used earlier in the process alongside or in place of traditional culture methods.

Running multiple systems in parallel (culture and molecular) is viewed as expensive and inefficient by many laboratory managers, who are reluctant to adopt molecular methods exclusively for a variety of reasons. One of those reasons is a lack of confidence in molecular approaches due to a paucity of robust data proving the equivalence (or better) of molecular approaches compared with culture approaches.

To explore these issues further a workshop was organised by the FSA in January 2012 entitled “*The Application of Molecular Epidemiology to Investigations of Foodborne Disease Outbreaks: Current Status and Future Plans*”. (http://www.food.gov.uk/multimedia/pdfs/ngsworkshop.pdf) This workshop brought together key experts and stakeholders to discuss the opportunities and challenges presented by emerging technologies. There was enthusiasm in the workshop for the potential of the newer generations of sequencing tools, which were widely believed to be capable of providing improved information over currently used approaches at a comparable or lower cost. Clear examples of where such approaches could

immediately add significant value if used in an outbreak situation were described which aligned closely with the areas discussed in this review.

Discussions around the practical steps required to facilitate efficient, appropriate and timely transfer of the technologies from high-throughput academic centres of excellence to reference laboratories and eventually to front-line laboratories identified hurdles that needed to be overcome. Issues such as staff training, data handling, storage and interpretation, availability of reference databases, backwards comparability to historical datasets and quality systems likely to be required were covered.

The workshop generated the following key conclusions and recommendations:

- High-throughput sequencing is currently capable of providing a significant benefit to outbreak investigations and should be used from the next outbreak onwards

- Whilst technical, quality, logistical and training issues need to be addressed these issues should not delay initiation of the roll-out of the technology

- Roll-out is expected to be in phases, with transfer from academic centres of excellence where the technologies are already established to key “early-adopter” clinical and reference laboratories occurring immediately, transfer to all reference laboratories within the next 2-5 years and to all front-line laboratories (clinical, epidemiological, food, animal) within the next 5-10 years

- An audit of current methods should be undertaken to help the Agency understand how soon such techniques could be widely replaced by sequencing

- Efficient implementation will require co-operation between multiple funding bodies (FSA, DEFRA, DH and funding councils) and funding should be allocated to aid the transition

- Consideration and implementation of standardisation as sequencing technologies become more widely used will be crucial. Standardisation of methods should be considered where possible, alongside clear method performance assessments to identify sources of variability and uncertainty between the different methodologies

- It should be possible to train current laboratory staff to generate the data, but analysis of the data will require further specialist training and expertise

- Interpretation software that is widely accepted, easy to understand and interpret needs to be developed

- There is a current lack of well-curated, inter-operable and quality controlled databases containing microbial sequences. Examples of where such databases do exist include a DEFRA/FSA supported campylobacter database (pubmlst.org/campylobacter). Funding should be allocated to help build and maintain wider pathogen sequence databases and to sequence historical isolates to help populate the databases

- Standardisation in the recording of additional information is also needed. It is clear that to be of maximum benefit genome sequence information needs to be integrated with other information from environmental, clinical and animal data and this additional information needs to be collected at the time and not retrospectively

- The Agency needs to engage with those developing policy at the clinical end (Public Health England) to help reduce the timescales and improve consistency of testing and reporting

- The gaps in knowledge in animal populations also need to be considered and funding should be allocated to look at endemic levels of zoonotic pathogens

In an outbreak scenario, the accuracy of identification must be at least as reliable as standard culture/serotyping methods. The consequence of a false identification could lead to more people becoming ill as the source is not found and economic losses for any producers of wrongly implicated foodstuffs. Therefore, generation of data to directly compare and confirm the reliability of molecular approaches in a comprehensive and valid manner will be a first step to wider acceptance.

11. Conclusion

This review has discussed the potential of current and emerging molecular approaches to improve the foodborne disease outbreak investigation process. There is a substantial amount of evidence in the literature demonstrating the benefits of using molecular approaches and the technologies are evolving at such a fast pace that further improvement in performance, speed, throughput and costs will continue to occur.

What is currently lacking and hindering uptake of molecular approaches is an appropriate framework to facilitate validation and implementation of these newer approaches (especially NGS) outside of large academic centres. A review of current practice should be undertaken in order to facilitate adoption of molecular approaches, reduce the burden on individual laboratories and harmonise approval of novel tests in line with industrial guidelines.

Despite the promise of molecular methods over conventional techniques, there are few studies that have comprehensively made this comparison directly. Molecular based assays typically detect the genomic DNA from pathogens which may be present whether the organism is viable or not, whereas culture based approaches will only detect viable organisms that can be cultured in a laboratory. The paucity of data regarding the increased number of positive samples found with nucleic acid analysis, and the implications of this, will be a major obstacle to the adoption of these methods. Efforts to define better a positive result and demonstrate equivalence (or superiority) to culture based approaches will be required. More advanced molecular approaches to detect evidence of viability (e.g. gene expression) may also help to define better positive results.

The genomics, clinical, food, epidemiology, informatics and measurement communities will need to work together alongside the various funding bodies to help improve uptake by developing harmonised standards and reference

methods/materials for using these novel approaches to ensure data compatibility and platform comparability. Comprehensive inter-operable reference databases and high quality analysis tools that are user friendly and generate easy to understand outputs for use by non-informatics personnel will also need to be developed in parallel.

12. Acknowledgements

We are grateful to the FSA for funding this work. We are also grateful to all those who have provided expert help and advice during the course of this project.

13. References

Abubakar, I., L. Irvine, et al. (2007). "A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food." Health Technol Assess 11(36): 1-216.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol 215(3): 403-410.

Amar, C. F., C. L. East, et al. (2007). "Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996)." Eur J Clin Microbiol Infect Dis 26(5): 311-323.

Anjum, M. F., M. Mafura, et al. (2007). "Pathotyping *Escherichia coli* by using miniaturized DNA microarrays." Appl Environ Microbiol 73(17): 5692-5697.

Annonymous (2011). "Outbreak genomics." Nat Biotech 29(9): 769-769.

Asiello, P. J. a. B., A.J. (2011). "Minaturized isothermal nucliec acid amplification, A review." Lab Chip. 11(8): 1420-1430.

Bai, S., J. Zhao, et al. (2010). "Rapid and reliable detection of 11 food-borne pathogens using thin-film biosensor chips." Appl Microbiol Biotechnol 86(3): 983-990.

Barrett T.J. , P. G.-S., and B. Swaminathan (2006). "Interpretation of Pulsed-Field Gel Electrophoresis Patterns in Foodborne Disease Investigations and Surveillance." Foodborne Pathogens and Disease. 3(1): 20-31. .

Boxrud, D. (2010). "Advances in subtyping methods of foodborne disease pathogens." Curr Opin Biotechnol 21(2): 137-141.

Buchholz, U., H. Bernard, et al. (2011). "German outbreak of *Escherichia coli* O104:H4 associated with sprouts." N Engl J Med 365(19): 1763-1770.

Bugarel, M., L. Beutin, et al. (2011). "Identification of genetic markers for differentiation of Shiga toxin-producing, enteropathogenic, and avirulent strains of *Escherichia coli* O26." Appl Environ Microbiol 77(7): 2275-2281.

Campden BRI (2012). Catalogue of rapid microbiological methods - 7th Ed 2012 (Review 1) Campden BRI, Campden.

Caporaso, J. G., J. Kuczynski, et al. (2010). "QIIME allows analysis of high-throughput community sequencing data." Nat Methods 7(5): 335-336.

Casey PG, H. C., Gahan CG. (2011). "E. coli O104:H4: social media and the characterization of an emerging pathogen." Bioeng Bugs. 2(4): 189-193

Cherkaoui, A., Hibbs, J., Emonet, S., Tangomo, M., Girard, M., Francois, P., and Schrenzel, J. (2010.). "Comparison of Two Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Methods with Conventional Phenotypic Identification for Routine Identification of Bacteria to the Species Level. ." J. Clin. Microbiol. 48(4): 1169-1175.

Chin, C. S., J. Sorenson, et al. (2011). "The origin of the Haitian cholera outbreak strain." N Engl J Med 364(1): 33-42.

Clarridge, J. E., 3rd (2004). "Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases." Clin Microbiol Rev 17(4): 840-862, table of contents.

Cole, J. R., Q. Wang, et al. (2009). "The Ribosomal Database Project: improved alignments and new tools for rRNA analysis." Nucleic Acids Res 37(Database issue): D141-145.

Consortium, H. M. P. (2012). "Structure, function and diversity of the healthy human microbiome." Nature 486(7402): 207-214.

Devonshire AS, S. R., Wilkes TM, Taylor MS , Foy CA, Huggett JF (2012). "Application of next generation qPCR and sequencing platforms to mRNA biomarker analysis." Methods In press.

Dohm, J. C., C. Lottaz, et al. (2008). "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing." Nucleic Acids Res. 36(16): e105.

Dohm, J. C., C. Lottaz, et al. (2008). "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing." Nucleic Acids Res 36(16): e105.

Fitzgerald, C., M. Collins, et al. (2007). "Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups." J Clin Microbiol 45(10): 3323-3334.

Fleischmann RD, A. M., White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science. 269(5223): 496-512.

Fukushima, H., J. Kawase, et al. (2010). "Simultaneous Screening of 24 Target Genes of Foodborne Pathogens in 35 Foodborne Outbreaks Using Multiplex Real-Time SYBR Green PCR Analysis." Int J Microbiol 2010.

Geue, L., S. Schares, et al. (2010). "Rapid microarray-based genotyping of enterohemorrhagic *Escherichia coli* serotype O156:H25/H-/Hnt isolates from cattle and clonal relationship analysis." Appl Environ Microbiol 76(16): 5510-5519.

Gill, P. a. G., A. (2008). "Nucleic Acid isothermal amplification technologies: A Review. ." Nucleosides Nucleotides and Nucleic Acids. 27(3): 224-243.

Gonzales, T. K., M. Kulow, et al. (2011). "A high-throughput open-array qPCR gene panel to identify, virulotype, and subtype O157 and non-O157 enterohemorrhagic Escherichia coli." Mol Cell Probes 25(5-6): 222-230.

Haas, B. J., D. Gevers, et al. (2011). "Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons." Genome Res 21(3): 494-504.

Hannis, J. C., S. M. Manalili, et al. (2008). "High-resolution genotyping of Campylobacter species by use of PCR and high-throughput mass spectrometry." J Clin Microbiol 46(4): 1220-1225.

Havelaar, A. H., S. Brul, et al. (2010). "Future challenges to microbial food safety." Int J Food Microbiol 139 Suppl 1: S79-94.

Hockin, J. C., J. Y. D'Aoust, et al. (1989). "An international outbreak of *Salmonella nima* from imported chocolate." Journal of Food Protection 52(1): 51-54.

HPA (2011). Working towards a Next Generation Sequencing Strategy across the HPA. H. P. Agency. London, Health Protection Agency, London,.

HPA (2012). UK Standards for Microbiology Investigations - Commercial and In-House Diagnostic Tests: Evaluations and Validations, Standards Unit, Microbiology Services Division, Health Protection Agency, London. Q 1, Issue no: 4.2.

Huang, Q., Q. Hu, et al. (2007). "Identification of 8 foodborne pathogens by multicolor combinational probe coding technology in a single real-time PCR." Clin Chem 53(10): 1741-1748.

Jolley, K. A., M. S. Chan, et al. (2004). "mlstdbNet - distributed multi-locus sequence typing (MLST) databases." BMC Bioinformatics 5: 86.

Jones, T. F. and P. Gerner-Smidt (2012). "Nonculture diagnostic tests for enteric diseases." Emerg Infect Dis 18(3): 513-514.

Josefsen, M. H., C. Löfström C, T.B. Hansen, L.S. Christensen, J.E. Olsen, and J. Hoorfar (2010). "Rapid quantification of viable Campylobacter bacteria on chicken carcasses, using real-time PCR and propidium monoazide treatment, as a tool for quantitative risk assessment. ." Appl. Environ Microbiol 76(5097-5104.).

Kort, R., B. J. Keijser, et al. (2008). "Transcriptional activity around bacterial cell death reveals molecular biomarkers for cell viability." BMC Genomics 9: 590.

Kupferschmidt, K. (2011). "Epidemiology. Outbreak detectives embrace the genome era." Science 333(6051): 1818-1819.

Leader, B. T., J. G. Frye, et al. (2009). "High-throughput molecular determination of salmonella enterica serovars by use of multiplex PCR and capillary electrophoresis analysis." J Clin Microbiol 47(5): 1290-1299.

Lehmann, L. E., K. P. Hunfeld, et al. (2008). "A multiplex real-time PCR assay for rapid detection and differentiation of 25 bacterial and fungal pathogens from whole blood samples." Med Microbiol Immunol 197(3): 313-324.

Lienau, E. K., E. Strain, et al. (2011). "Identification of a salmonellosis outbreak by means of molecular sequencing." N Engl J Med 364(10): 981-982.

Maiden, M. C., J. A. Bygraves, et al. (1998). "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms." Proc Natl Acad Sci U S A 95(6): 3140-3145.

Malecki, M., V. Schildgen, et al. (2012). "Rapid screening method for multiple gastroenteric pathogens also detects novel enterohemorrhagic Escherichia coli O104:H4." Am J Infect Control 40(1): 82-83.

Mazzeo, M. F., A. Sorrentino, et al. (2006). "Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the discrimination of food-borne microorganisms." Appl Environ Microbiol 72(2): 1180-1189.

McDonald, D., M. N. Price, et al. (2012). "An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea." ISME J 6(3): 610-618.

Mellmann, A., D. Harmsen, et al. (2011). "Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology." PLoS ONE 6(7): e22751.

Mitra, S., M. Stark, et al. (2011). "Analysis of 16S rRNA environmental sequences using MEGAN." BMC Genomics 12 Suppl 3: S17.

Mody, R. K., S. A. Greene, et al. (2011). "National outbreak of Salmonella serotype saintpaul infections: importance of Texas restaurant investigations in implicating jalapeno peppers." PLoS ONE 6(2): e16579.

Mullis K, F. F., Scharf S, Saiki R, Horn G, Erlich H. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol 51(Pt 1): 263-273.

Notomi T, O. H., Masubuchi H, Yonekawa T, Watanabe K, Amino N, Hase T. (2000). "Loop-mediated isothermal amplification of DNA." Nucleic Acids Res. 28(12): E63.

O'Brien, S. (2012). "The public health impact of food-related illness." Curr Opin Infect Dis. In press.

Perkel, J. (2011). "Making contact with sequencing's fourth generation." Biotechniques 50(2): 93-95.

Petrosino, J. F., S. Highlander, et al. (2009). "Metagenomic pyrosequencing and microbial identification." Clin Chem 55(5): 856-866.

Phillips, G., B. Lopman, et al. (2009). "Diagnosing norovirus-associated infectious intestinal disease using viral load." BMC Infect Dis 9: 63.

Phillips, G., B. Lopman, et al. (2009). "Diagnosing rotavirus A associated IID: Using ELISA to identify a cut-off for real time RT-PCR." J Clin Virol 44(3): 242-245.

Pires, S. M., E. G. Evers, et al. (2009). "Attributing the human disease burden of foodborne infections to specific sources." Foodborne Pathog Dis 6(4): 417-424.

Pruesse, E., C. Quast, et al. (2007). "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB." Nucleic Acids Res 35(21): 7188-7196.

Rasko, D. A., D. R. Webster, et al. (2011). "Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany." N Engl J Med 365(8): 709-717.

Richmond, G. S., H. Khine, et al. (2011). "MassCode liquid arrays as a tool for multiplexed high-throughput genetic profiling." PLoS ONE 6(4): e18967.

Rohde, H., J. Qin, et al. (2011). "Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4." N Engl J Med 365(8): 718-724.

Rothberg, J. M., W. Hinz, et al. (2011). "An integrated semiconductor device enabling non-optical genome sequencing." Nature 475(7356): 348-352.

Saunders, N., I. Sharp, et al. (2011). Guidance on the Development and Validation of Diagnostic Tests that Depend on Nucleic Acid Amplification and Detection, Health Protection Agency.

Singh, P., S. L. Foley, et al. (2012). "Multilocus sequence typing of Salmonella strains by high-throughput sequencing of selectively amplified target genes." J Microbiol Methods 88(1): 127-133.

Suzuki, S., N. Ono, et al. (2011). "Comparison of Sequence Reads Obtained from Three Next-Generation Sequencing Platforms." PLoS ONE 6(5): e19534.

Tam, C. C., S. J. O'Brien, et al. (2012). "Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease." Clin Infect Dis 54(9): 1275-1286.

Tankouo-Sandjong, B., A. Sessitsch, et al. (2008). "Development of an oligonucleotide microarray method for Salmonella serotyping." Microb Biotechnol 1(6): 513-522.

Thompson, S. A. (2008). Feeling the Squeeze - Tomato growers take big hit in food scare. Rural Cooperatives, Rural Development, Business and Cooperative Programs, U.S. Department of Agriculture, 1400 Independence Avenue, SW, Stop 0705, Washington, DC.: 13-15, 41.

Williams, N. K. E. a. C. M. (2007). Infectious Disease Epidemiology: Theory And Practice 2nd edition Burlington, MA. , Jones & Bartlett Learning.

14. Appendix 1: Brief overview of current molecular approaches

14.1. Polymerase Chain Reaction (PCR)

Standard PCR methods typically utilise short oligonucleotide primers (typically 18-25 bases long) specific for target sequences and a *Taq* polymerase enzyme to amplify the DNA target region in a cycling temperature assay. The resulting PCR product (amplicon) is then subjected to agarose gel electrophoresis for visualisation. A visible band of the correct size indicates a positive sample.

PCR based approaches typically do not require a pure culture of the isolate under investigation as the specificity of the process is such that target sequences can be amplified and detected in the presence of large amounts of non-target sequences (e.g. from food or clinical matrices or in the presence of other microbes). Therefore, the time to detection can be reduced and organisms which are difficult to cultivate or are non-culturable can be identified. Some methods may benefit from enrichment prior to PCR (as discussed below) to reduce the effect of inhibitors and to increase sensitivity. Whilst the use of enrichment adds to the total length of the test, it usually only an additional 24-48 hours and this is still an improvement over traditional methods which require enrichment, subculture to selective agar and isolate identification which can take 5-7 days. PCR methods require DNA extraction which can be performed using commercially available kits to make the process faster and more amenable to automation.

PCR primers can be designed to target universal bacterial genes such as the 16S rRNA subunit gene which is often used for speciation (Clarridge 2004). This gene is comprised of conserved regions to which universal primers are designed to target and variable regions which are used for comparative taxonomy. PCR primers can also be designed to be specific for genes involved in virulence, antigeniticy and regulation. This genetic information can provide an organism's serotype which can help to determine the subtype.

All PCR methods require appropriate controls as false negative results can occur due to the inhibitory effects of the matrix which is present in nucleic acid extractions from food and clinical samples. Controls for false positive results also need to be included to militate against contamination and specificity issues. Furthermore, PCR based approaches will detect DNA targets present in a sample regardless of whether they originate from a viable or non-viable organism. Therefore, if proof of viability is needed an initial (or parallel) culture step will be required. Alternatively, PCR assays designed to target viability markers such as gene expression transcripts can be used.

14.2. Pre-PCR enrichment approaches

In contrast to clinical samples, food samples may have a very low pathogen load (e.g. in chocolate, <10cfu/g *Salmonella* is thought to be enough to cause illness (Hockin, D'Aoust et al. 1989)). This potential low loading could lead to a false-negative result if steps are not taken to ensure that the number of target organisms applied to a particular detection system is not greater that the limit of detection of that system. To guard against this, the level of target microorganisms from food samples usually has to be increased either through encouraging them to grow

(usually through incubation in suitable enrichment broth) or through a method that concentrates the target organism that can be detected using the method chosen, or through a combination of both of these approaches

Enrichment often requires the incubation of the sample in a nutrient broth for several hours to encourage growth of the target organism. Therefore, most assays still require up to 24h from time of sample submission to a positive/negative result. This is an improvement over current culture-based methodology which would need one to two stages of enrichment growth in a broth medium followed by detection on a diagnostic agar. This could take at least 48h and perhaps up to 120h to provide an initial positive/negative result.

The ease of application, and the low cost of enrichment via culture have led to it being widely adopted as the first step in the vast majority of the PCR based applications for the food industry. Other methods rely on immuno-magnetic separation (IMS), the principle of which is outlined in Figure 5.

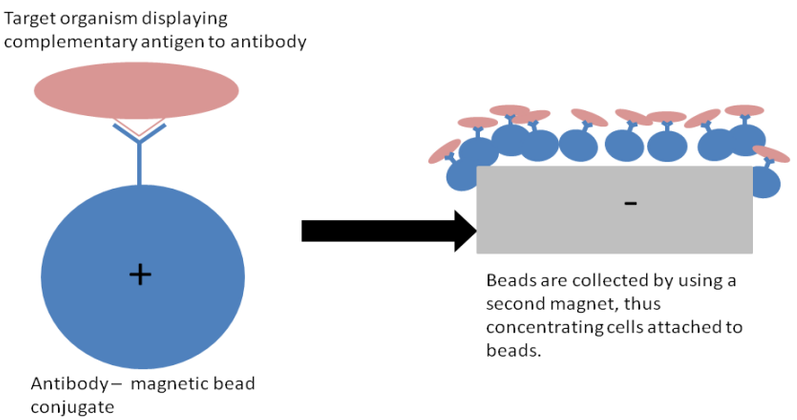


FIGURE 23 OVERVIEW OF IMMUNO-MAGNETIC SEPARATION

Generally, samples are blended in some form of diluent and mixed with the bead-antibody conjugate. The mixture is then allowed to stand for a short period of time to allow for the beads and their targets to bind. The mixture of sample and beads is then passed over a magnet to pull the beads and their bound targets out of the mixture, which is then discarded. The beads can then be washed and used in PCR reactions. This method of enrichment has the advantage of being a swift way to concentrate cells, taking minutes rather than hours compared with culture based methods. A further advantage is that the resulting bead/bacteria conjugate will be much cleaner and so subsequent enzymatic reactions (e.g. in PCR) will be more efficient through less carryover of contaminating solutes from the sample matrix.

Whist IMS can be used without the need for an enrichment/growth step, in many instances a short growth period is used to slightly increase target cell numbers before IMS is applied. IMS is used as a part of some standard reference methodologies e.g. BS EN ISO 16654 for the detection of *E.coli* O157.

14.3. Sequencing

Standard Sanger sequencing relies on incorporating labelled dideoxy nucleotide analogues (ddNTPs) that terminate the synthesis of DNA complementary to the

sequence to be read. If these analogues are used in conjunction with standard deoxy nucleotides (dNTPs) which are incorporated in the usual way and allow subsequent binding of further nucleotides, then the final reaction volume will contain a mixture of fragments that differ from each other in length by one base pair, with the terminal pair labelled either by fluorescence or radioactivity (today fluorescence is used in the overwhelming majority of reactions). As different labels can be attached to each ddNTP (ddATP, ddGTP, ddCTP and ddTTP), the sequence is determined by separating the DNA molecules by size and identifying the label on each set of molecules of the same size. Acrylamide gel based methods followed by manual identification of labels were used for separation originally, with column based capillary electrophoresis separation and automated identification taking over as technology improved. Sanger sequencing is used in the food industry to sequence variable regions of genes coding for ribosomal RNA (rRNA) subunits in bacteria, moulds and yeasts. Its use in clinical applications is to aid identification of similarities between outbreak strains and to aid with identification of isolates to a deeper level than species.

14.4. Pulsed Field Gel Electrophoresis (PFGE)

PFGE requires extraction of DNA from a pure culture and involves a restriction digestion of DNA (using an enzyme with numerous recognition sites along the genome of interest) to generate a characteristic range of DNA fragments varying in length, which are separated and resolved during electrophoresis. As the name suggests fragment resolution is enhanced, by applying pulses of current across the gel during electrophoresis. This technique is recognised as the gold standard approach for the characterisation of bacterial pathogens (particularly for epidemiological investigations) and to assist rapid dissemination of data, international and national databases have been established such as SalmGene (http://www.hpa-bionum.org.uk/bionumerics/salm_gene/ Developed by the HPA and funded by the European Commission - DG Research) and PulseNet (http://www.cdc.gov/pulsenet/ - developed by CDC in association with the Association of Public Health Laboratories APHL) respectively.

In an outbreak situation, this technique is used to compare strains that have caused illness, and is often referred to as ‘Molecular Fingerprinting’ due to the ability to differentiate between strains at sub species level that cannot be discriminated using biochemical approaches.

There are a number of drawbacks to PFGE and these include the level of skill required to achieve consistency of analysis, time to result, and the hands on time required compared to other molecular based tests due a lack of method automation. Pattern interpretation requires manual input to take into consideration method reproducibility, gel quality, genetic variability of the species being sub-typed and the prevalence of the profile being analysed (Barrett T.J. 2006). The dependency on suitable restriction digestion sites to discriminate similar strains also limits the ability to fully identify all potential strain variants.

14.5. Ribotyping

Ribotyping relies on the fact that all organisms (other than viruses) possess at least one gene coding for particular sub-units of the ribosome (rDNA). These genes

contain some highly conserved sequences across bacterial species, and between these conserved sequences there are regions of DNA that have mutated at a relatively constant rate over time. Ribotyping uses probes that target the highly conserved ribosomal regions to detect the areas of variability within these sequences on the bacterial genome. The technique starts with a restriction digest of the genomic DNA as described in the PFGE section above to generate a distinct spectrum of fragments. These fragments are separated according to size through a gel matrix, and transferred onto a membrane for detection using labelled probes specific to targets within the ribosomal genes (typically 16S rDNA and 23S rDNA). It is used in the food industry to compare strains of bacteria, as it does have the ability to discriminate between not only different species, but different strains within species. This technique has been automated through the introduction of the RiboPrinter that is able to give results within 9h from a pure colony. Automation allows standardisation of both the technique and its results, across user laboratories including defined thresholds for analysing isolate similarity (within the software). Additional benefits of standardisation are the ability to develop comprehensive database of patterns that enable comparison of patterns obtained today from those obtained from past analyses.

Disadvantages include the relatively high cost and time-consuming nature of the method, as well as the reliance on a culture step and suitable restriction digestion sites for discrimination (as discussed above for the PFGE approach). This technique focuses on a defined set of genes which may limit the level of differentiation at a sub species level in comparison to other methods discussed in this review that offer a more genome-wide approach.

14.6. Multilocus sequence typing (MLST)

Multilocus sequence typing (MLST) utilises a panel of genes and compares variability of the nucleotide sequence data within those genes (approximately 450-500 bp in length) for each target to assess genetic similarity. This technology is well developed for bacterial isolates (Maiden, Bygraves et al. 1998), and has allowed the application of standardised primer sets (published on web based databases such as PubMLST hosted at the University of Oxford, UK (Jolley, Chan et al. 2004), which facilitates inter-laboratory data comparison for epidemiological investigations.

The technique relies on sequencing a suite of genes for each organism; therefore the time from sample submission to identification via MLST is too long for it to be used as the primary identification method, despite its accuracy. Instead, it is used to confirm similarity between isolates at a sub species level. Advantages of MLST are the objectivity of the analysis which focuses on genetic sequences rather than DNA fragment profiles which lends itself more easily to inter-laboratory comparisons. This technique does however have an intolerance to sequence errors that may influence the typing profile obtained (Williams 2007).

Newer sequencing approaches discussed in later sections of this review are cost-effectively enabling the analysis of entire genomes rather than just a few select genes. The improved level of information gained from whole genome sequencing is enabling greater insights in to strain differentiation and identification and is likely to eventually replace standard MLST typing.

14.7. Variable Number Tandem Repeat (VNTR)

VNTR analysis is a PCR-based method which determines the number of small repeat regions occurring along a particular gene, which are detected through variations in amplimer length. This technique was originally applied to human DNA fingerprinting (Cawood, 1989), and has been transferred to microbial analysis with a focus on clinically relevant organisms such as methicillin resistant Staphylococcus aureus (MRSA) and Mycobacterium tuberculosis (Frothingham and Meeker O'Connell 1998, Sabat et al. 2003).

In recent years VNTR has been expanded to encompass multiple gene targets in multiple locus VNTR analysis (MLVA), to increase the power to differentiate between closely related isolates. Internet resources for the development of bacterial assays are available (Deneoud and Vergnaud, 2004), along with another broad range package that is able to predict tandem repeats in eukaryotes (Legendre et al, 2007). This technique is currently being used in the United States at the PulseNet Next Generation Subtyping Methods Unit as a complimentary method to PFGE in two areas:

- select multistate clusters of strains as prioritised by CDC PulseNetLab
- special investigations by CDC, USDA, FDA, and local / state epidemiologists

There are limited databases for MLVA profiles at present with information available for E. coli O157, S. Typhimurium and S. Enteritidis. MLVA has been reported to be a very useful tool in outbreak investigations with the ability to separate outbreak cases from background related cases as well as determining the relatedness of PFGE patterns sharing a high degree of similarity (Sabol, 2012)

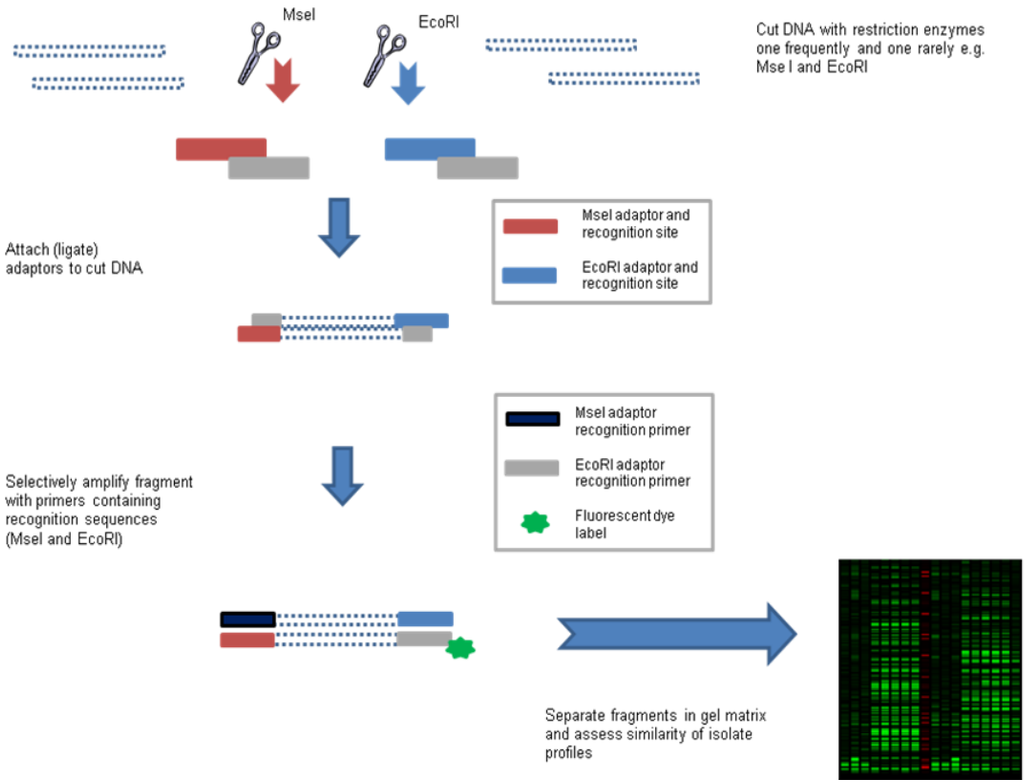
One of the key drawbacks to this technique is the development of an effective set of loci to enable differentiation at subspecies level. Although the technique may be reproducible in the same laboratory, standardisation of the method is essential for inter laboratory studies, to compensate for differences in equipment and gel variability used to determine fragment length. Standardized MLVA protocols are currently available at PulseNet* (for STEC O157 and S. enterica serotype Typhimurium) however they are platform specific. The introduction of other platforms and organisms would be require validation prior to their release to laboratories.

*Available at: http://www.pulsenetinternational.org/protocols/Pages/mlva.aspx

14.8. Fluorescent amplified-fragment length polymorphism (fAFLP)

fAFLP is a multi-stage analysis used to assess strain similarity of pure cultures. The technique involves enzymatic double digestion of the total genomic DNA, followed by the ligation of an adaptor which is compatible to the resulting overlapping or ‘sticky’ ends of the DNA fragments. These restriction fragments are simultaneously amplified and labelled during a PCR with primers designed to bind to the adaptor sites (attached to the fragments in the ligation step). One of the primers used in the PCR step is labelled with a fluorescent dye to add a ‘colour tag’ to the amplicons

which enables fragment detection during electrophoresis on a sequencer. A schematic detailing the steps involved in fAFLP is shown below.



Schematic of Amplified Fragment Length Polymorphism (AFLP) analysis

The use of fluorescently labelled primers to assist in fragment detection has enabled increased throughput using automated sequencer systems. This technique has been reported to provide a greater resolution and be less sensitive to DNA quality than PFGE (Zhao et al 2000). Further research has indicated that fAFLP is less sensitive to the genetic instability of Campylobacter that is known to be problematic to other typing methods (Wassamer and Newell, 2000).

The use of capillary based electrophoresis for this method has significantly improved the ability of the method to provide rapid, robust results. Benefits of such systems include the ability to run the internal sizing standard simultaneously with the sample, and the capacity to detect multiple dyes thus enabling 2 digests to be mixed immediately prior to loading onto the capillary (Lindstedt et al 2000). Sample run times on capillary based systems are relatively rapid with certain platforms able to detect +/- 1bp differences in fragments in 30 minutes (Lindstedt et al 2000). The fast run time coupled with the availability of multi-capillary platforms gives the potential to scale up to very high throughput. An additional advantage of AFLP is its potential to determine genetic markers with epidemiological significance that could be used in subsequent outbreak investigations (Hopkins et al 2004, Tamada et al 2001).

Although AFLP is a very powerful tool; the set up cost, additional consumables and high level of skill required limits its potential use in routine laboratories. The technique is a proprietary technology owned by Keygene and a licence is required

for its use in service laboratories. Further details of the licence requirements are available on the Keygene website at http://www.keygene.com/services/licenses.php

15. Appendix 2: Commercially available PCR assays for foodborne pathogens in food or environmental samples

Organism	Name of kit	Company	Type of PCR	Claimed analysis time	Validating bodies
Salmonella spp.	Adiafood	AES Chemunex	Real-Time	19h	ISO 16140
Salmonella enteriditis	TaqMan	Life Technologies	qPCR	27h	FDA
Salmonella spp	MicroSEQ	Life Technologies	qPCR	19h	AOAC
Salmonella II	iQ-Check	Bio-Rad	qPCR	24h	NF, AOAC-RI, NordVal
Salmonella spp.	Foodproof	Biotecon	Real-Time	24h	AOAC, NordVal, Microval
Salmonella spp	BAX	DuPont Qualicon	Real-Time	24h	AOAC, USDA, AFNOR, NPIP, Health Canada MFLP-29, Brazil MAPA Directive 41, NordVal, Danish Vetinary and Food Administration.
Salmonella spp.	R.A.P.I.D. LT Food Security System	Idaho Technology Inc	Real-Time	17h	AOAC
Salmonella spp. (and *Salmonella / Escherichia coli* O157)	GeneDisc	Pall	Real-Time	10h or 18h	AFNOR, AOAC
Salmonella spp.	QuickBlue	Q Bioanalytic	Real-Time	5h post-enrichment	N/A
Salmonella spp. *Salmonella spp.*	Mericon 3M	QIAGEN Ltd 3M	Real-Time LAMP Isothermal Amplification, Bioluminescent detection	3h post-enrichment 1h sfter enrichment	N/A (pending AOAC and AFNOR) Pending AOAC RI

Salmonella spp.	SAS Molecular Tests	SAS Scientific	LAMP Isothermal Amplification	1h after enrichment	AOAC RI
Listeria monocytogenes	Adiafood	AES Chenunex	Real-Time	<24h	AOAC, CFIA
L. monocytogenes	TaqMan	Life Technologies	Real-Time	28h	AOAC
Listeria spp.	MicroSEQ	Life Technologies	Real-Time	31h	N/A
L. monocytogenes	Assurance GDS	BioControl Systems Inc	Real-Time	24h+	AOAC
Listeria spp.	Assurance GDS	BioControl Systems Inc	Real-Time	32h	AOAC
L. monocytogenes II	iQ-Check	Bio-Rad	Real-Time	24h	NF, AOAC-RI, NordVal
Listeria spp.	iQ-Check	Bio-Rad	Real-Time	24h	NF, AOAC-RI,
Listeria spp.	Foodproof	Biotecon	Real-Time	48h	N/A
L. monocytogenes	Foodproof	Biotecon	Real-Time	48h	AOAC-RI, NordVal
L. monocytogenes 24E	BAX	Dupont Qualicon	Real Time	24 – 48h	AOAC, AOAC-RI, Health Canada MFLP 28, Swedish National Food Administration, Brazil MAPA MLG 8A.01
Listeria 24E	BAX	Dupont Qualicon	Real Time	24 – 48h	AOAC-RI, Health Canada MFLP-15e, Swedish National Food Administration, Brazil MAPA MLG 8A.01
L.monocytogenes	BAX	Dupont Qualicon	Real Time	24h	AOAC, AOAC-RI, Health Canada MFLP 28, Swedish National Food Administration,

					Brazil MAPA MLG 8A.01
Listeria spp.	BAX	Dupont Qualicon	Real Time	24h	AOAC-RI, Health Canada MFLP-15e
Listeria spp.	BAX	Dupont Qualicon	Real Time (Reverse transcriptase assay)	8h (environmental samples)	AOAC-RI
Listeria spp.	R.A.P.I.D. LT Food Security System	Idaho Technology Inc	Real-Time	24h	N/A
L. monocytogenes	GeneDisc	Pall	Real-Time	19h	AFNOR
Listeria spp.	GeneDisc	Pall	Real-Time	19h	AFNOR
Listeria DUO	GeneDisc	Pall	Real-Time	19h	AFNOR
Listeria spp	*Listeria* Capture Kit	Profos AG	Real-Time	21h	AOAC-RI
L. monocytogenes	QuickBlue	Q Bioanalytic	Real-Time	5h post-enrichment	N/A
Listeria innocua	QuickBlue	Q Bioanalytic	Real-Time	5h post-enrichment	N/A
Listeria Triple	QuickBlue	Q Bioanalytic	Real-Time	5h post-enrichment	N/A
Listeria spp	QuickBlue	Q Bioanalytic	Real-Time	5h post-enrichment	N/A
Listeria spp	Mericon	QIAGEN Ltd	Real-Time	3h post-enrichment	N/A
L. monocytogenes	Mericon	QIAGEN Ltd	Real-Time	3h post-enrichment	N/A
Listeria spp *Listeria spp.*	SureFood 3M Molecular Detection Assay. *Listeria*	R-Biopharm Ltd 3M	Real-Time LAMP Isothermal Amplification, Bioluminescent detection	24h 1h after enrichment	N/A

Escherichia coli O157	Adiafood	AES Chemunex	Real-Time	<24h	AOAC
E. coli O157:H7	Adiafood	AES Chemunex	Real-Time	<24h	AOAC, CFIA
E. coli O157:H7	TaqMan	Life Technologies	Real-Time	20h	AOAC, AFNOR
E. coli O157:H7	iQ-Check	Bio-Rad	qPCR	24h	AFNOR, AOAC-RI
E. coli O157:H7	Foodproof	Biotecon	Real-Time	24h	AOAC-RI, NordVal
E. coli O104	Foodproof	Biotecon	Real-Time	24h	
E. coli O157	BAX	DuPont Qualicon	Real Time	24h	AOAC-RI, Health Canada
E. coli O157:H7 MP	BAX	DuPont Qualicon	Real Time	8-24h	AOAC-RI
E. coli O157:H7	R.A.P.I.D. LT Food Security System	Idaho Technology Inc	Real-Time	40min post-enrichment	AOAC
E. coli stx1, stx2, eae assay	TaqMan	Life Technologies	Real-Time	24h	N/A
E. coli O104	TaqMan	Life Technologies	Real-Time	24h	N/A
E. coli O157:H7 (and *Salmonella & E. coli* O157)	GeneDisc	Pall	Real-Time	10h or 18h	AFNOR, AOAC
Shiga Toxin *E. coli* (and Shiga toxin *E. coli & Salmonella*)	GeneDisc	Pall	Real-Time	10h or 18h	AFNOR, AOAC
E. coli O157	*E. coli* O157 Capture kit	Profos-AG	Real-Time	6.5h	N/A
E.coli O157	SAS Molecular Tests	SAS Scientific	LAMP Isothermal Amplification	1h after enrichment	AOAC RI
E.coli O157	3M Molecular Detection	3M	LAMP Isothermal	1h after enrichment	

	Assay E.coli O157		Amplification, Bioluminescent detection		
EHEC	QuickBlue	Q Bioanalytic	Real-Time	5h post-enrichment	N/A
EHEC & EHECplus	GeneGen	Sy-Lab	Multiplex	24h post-enrichment	N/A
Staphylococcus aureus	Taqman	Life Technologies	Real-Time	24h	N/A
S. aureus	BAX	DuPont Qualicon	Real-Time	24-48h	AOAC-RI
S. aureus	QuickBlue	Q Bioanalytic	Real-Time	5h (post enrichment)	N/A
S. aureus	Mericon	QIAGEN Ltd	Real-Time	3h (post enrichment)	N/A
Campylobacter spp.	Adiafood	AES Chemunex	Real-Time	48h	AOAC
C. jejuni	TaqMan	Life Technologies	Real-Time	28h	N/A
Campylobacter spp. (detection)	Foodproof	Biotecon	Real-Time	24h	N/A
Campylobacter spp. (enumeration)	Foodproof	Biotecon	Real-Time	24h	N/A
Campylobacter spp.	iQ-Check	Bio-Rad	Real-Time	24h	N/A
C. jejuni &C. coli	BAX	DuPont Qualicon	Real-Time	24h	AOAC
Campylobacter spp.	Mericon	QIAGEN Ltd	Real-Time	3h (post enrichment)	N/A
C. jejuni, C. coli & C. laridis	Mericon	QIAGEN Ltd	Real-Time	3h (post enrichment)	N/A
Campylobacter spp.	SureFood	R-Biopharm Ltd	PCR and PCR-ELISA	24h	N/A
Bacillus cereus	*Bacillus cereus* detection	Biotecon	Real-Time	75min (post enrichment	N/A

	system			)	
B. cereus	QuickBlue	Q Bioanalytic	Real-Time	5h (post enrichment)	N/A
Cryptosporidium	R.A.P.I.D.® LT Water safety LT test kit	Idaho Technology	Real time	17h	None
Cryptosporidium	RT- PCR Detection Kit	Norgen Biotek Corp	Reverse transcript-se	3h for water samples	None
Cryptosporidium spp	KCRYP for Food & Environme-ntal samples	CEERAM tools	Real time	2.5h	None
Cryptosporidum parvum	PowerChek™ Cryptosporidum parvum Real-time PCR Kit	Kogene Biotech	Real time	3.5h post DNA extraction	None
Giardia intestinalis	RT- PCR Detection Kit	Norgen Biotek Corp	Reverse transcript-se	3h for water samples	None
Giardia spp	KGIAR for Food and Environmen-tal samples	CEERAM tools	Real time	2.5h	None
Giardia lamblia	PowerChek™ Giardia lamblia Real-time PCR Kit	Kogene Biotech	Real time	3.5h post DNA extraction	None
Vibrio parahaemolyticus,	QuickBlue	Q Bioanalytic	Real-Time	5h (post enrichment)	N/A
Vibrio p. kanagawa	QuickBlue	Q Bioanalytic	Real-Time	5h (post enrichment)	N/A
V. cholerae	QuickBlu	Q	Real-	5h (post enrichment	N/A

	e	Bioanalytic	Time	)	
V. alginolyticus	QuickBlue	Q Bioanalytic	Real-Time	5h (post enrichment)	N/A
V. vulnificus	QuickBlue	Q Bioanalytic	Real-Time	5h (post enrichment)	N/A
Vibrio spp.	SureFood	R-Biopharm Ltd	Real-Time	24h	N/A
Hepatitis A	KHAV for Food & Environme-ntal samples	CEERAM tools	Real time	2.5h post RNA extraction	None
Hepatitis E	KHEV for Food & Environmental samples	CEERAM tools	Real time	2.5h post RNA extraction	None
Norovirus GI	KNVGI for Food & Environmental samples	CEERAM tools	Real time	2.5h post RNA extraction	None

TABLE 3 PCR BASED KITS AVAILABLE TO THE FOOD INDUSTRY FOR RAPID DETECTION OF PATHOGENS.

Key: AOAC = Association of Analytical Communities, AOAC-RI = AOAC Research Institute, USDA = United States Department of Agriculture, AFNOR = Association Française de Normalisation, NPIP = National Poultry Improvement Plan (US), CFIA = Canadian Food Inspection Agency, MicroVal = European validation and certification organisation.

16. Appendix 3: Commonly used publicly available databases and search tools

BLAST (Basic Local Alignment Search Tool)

BLAST (Altschul, Gish et al. 1990) finds regions of local similarity between sequences. The program compares nucleotide sequences to sequence databases and calculates the statistical significance of matches. A recent addition to the BLAST programme is the development of a specific Microbial BLAST programme to focus on microbial taxonomy. (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=MicrobialGenomes)

SILVA

SILVA (Pruesse, Quast et al. 2007) (http://www.arb-silva.de/) provides datasets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) ribosomal RNA (rRNA) sequences for all three domains of life (Bacteria, Archaea and Eukarya).

Ribosomal database project (RDP)

The Ribosomal Database Project (RDP) (Cole, Wang et al. 2009) provides ribosome related data and services to the scientific community, including online data analysis and aligned and annotated Bacterial and Archaeal small-subunit 16S rRNA sequences. (http://rdp.cme.msu.edu/)

Greengenes

The Greengenes web application provides access to the 16S rRNA gene sequence alignment for browsing, blasting, probing, and downloading (McDonald, Price et al. 2012).

QIIME (Quantitative Insights Into Microbial Ecology)

QIIME (Caporaso, Kuczynski et al. 2010) is an open source software package for comparison and analysis of microbial communities, primarily based on high-throughput amplicon sequencing data. The software defines OTUs based on sequence similarity within the reads, and selects a representative sequence from each OTU. The OTU is assigned to a taxonomic identity using reference databases.

MEGAN

MEGAN (“MEtaGenome ANalyzer”) (Mitra, Stark et al. 2011) is widely used to perform the taxonomic and functional analysis of large metagenomic datasets. MEGAN simply assigns individual sequence reads to taxa by sequence similarity e.g. using BLAST without any initial clustering step.

GOLD (Genomes online database)

GOLD is a web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world. (http://www.genomesonline.org/cgi-bin/GOLD/index.cgi)

MG-RAST

The MG-RAST server is an open source system for annotation and comparative analysis of metagenomes. The server provides several methods for phylogenetic reconstructions. http://metagenomics.anl.gov/

17. Appendix 4: Review of expertise

The names and institutions in the following tables have been compiled from the following sources:

1. Attendees from the workshop organised by the FSA in January 2012 entitled "*The Application of Molecular Epidemiology to Investigations of Foodborne Disease Outbreaks: Current Status and Future Plans*". (http://www.food.gov.uk/multimedia/pdfs/ngsworkshop.pdf)
2. UK laboratories with NGS capability listed in the database maintained on omicsmaps.com
3. UK authors of relevant publications and conference presentations in the field
4. Names suggested during discussions with key UK stakeholders and experts in the field.

This list is not likely to be fully comprehensive and will need to be updated regularly.

UK Institutes with NGS capability (source: http://omicsmaps.com/)	Location (UK)
ARK-Genomics Roslin Institute	Edinburgh, Midlothian
Babraham Institute	Cambridgeshire
Barts Genome Centre	Greater London
BRC Genomics Facility, Guy's Hospital	Greater London
Cambridge Research Institute	Cambridgeshire
Cambridge Systems Biology Centre	Cambridge Systems Biology Centre
Centre for Genomic Research	Liverpool
Department of Biochemistry, University of Cambridge	Cambridgeshire
DSTL	Wiltshire
Health Protection Agency	Greater London
ICR	Greater London
Illumina (formerly Solexa HQ)	Essex
LGC Genomics	Greater London
Liverpool Cancer Research UK Centre, Liverpool, United Kingdom	Merseyside
London Research Institute	Greater London
MRC Clinical Sciences Centre	Greater London
MRC hub Cambridge (EASIH)	Cambridgeshire
NewGene	Newcastle Upon Tyne

Paterson Institute for Cancer Research	Manchester
Royal Brompton Hospital	Greater London
Sir Henry Wellcome Functional Genomics Facility (SHWFGF)	University of Glasgow, Glasgow
Source BioScience LifeSciences	Nottingham
Strathclyde Institute of Pharmacy and Biomedical Science	University of Strathclyde, Glasgow
Systems Biology Laboratory	Oxfordshire
The GenePool (University of Edinburgh)	City of Edinburgh
The Genome Analysis Centre	Norfolk
The MRC National Institute for Medical Research	London
The Sainsbury Laboratory	Norfolk
UCL Genomics	Greater London
University of Birmingham	Birmingham
University of Bristol	City of Bristol
University of Exeter	Devon
University of Liverpool Centre for Genomic Research	Liverpool
University of Manchester Centre for Integrated Genomic Medical Research	Manchester
University of Nottingham	Nottingham
University of Wales and Aberystwyth	Dyfed
University of York Technology Facility	York
Wales Gene Park	Cardiff
Warwick HRI	Coventry
Wellcome Trust Centre for Human Genetics	Oxfordshire
Wellcome Trust Sanger Institute	Cambridgeshire
West Midlands Regional Genetics Laboratory	Birmingham
Food/clinical microbiology/epidemiology experts	**Affiliation**
Dr Bob Adak	Health Protection Agency (HPA)
Dr Roy Betts	Campden BRI and Advisory Committee on Microbiological Safety of Food (ACMSF)
Prof Mark Blaxter	University of Edinburgh
Dr Derek Brown	Scottish Salmonella Reference Lab
Prof John Coia	ACMSF
Dr John Cowden	Health Protection Scotland (HPS)
Dr Lisa Crossman	The Genome Analysis Centre (TGAC)
Dr Richard Ellis	Animal Health and Veterinary Laboratories Agency

	(AHVLA)
Dr Ken Forbes	University of Aberdeen
Dr Carole Foy	LGC
Dr Vanya Gant	University College London Hospitals (UCLH)
Dr Kathie Grant	HPA
Dr Jonathan Green	HPA
Dr Rebecca Hodges	Medical Research Council (MRC)
Prof Rowland Kao	University of Glasgow
Prof Doug Kell	Biotechnology and Biological Sciences Research Council (BBSRC)
Asst Prof Mette Voldby Larsen	The Technical University of Denmark (DTU)
Prof Martin Maiden	University of Oxford
Prof Duncan Maskell	University of Cambridge
Prof Sarah O'Brien	ACMSF
Dr Julian Parkhill	The Sanger Centre
Dr Norval Strachan	University of Aberdeen
Dr Adam Staines	BBSRC
Prof John Wain	University of East Anglia (UEA)
Dr Alan Walker	The Sanger Centre
Prof Brendan Wren	London School of Hygiene and Tropical Medicine (LSHTM)
Prof Maria Zambon	HPA
Prof Mike Peck	Institute of Food Research (IFR)
Prof Neil Hall	University of Liverpool
Prof Martin Woodward	AHVLA
Dr Jim McLauchlin	HPA
Prof Andrew Fox	HPA
Prof John Threllfall	HPA
Dr Bert Popping	Eurofins Scientific Group, UK
Professor Bill Ollier	University of Manchester