

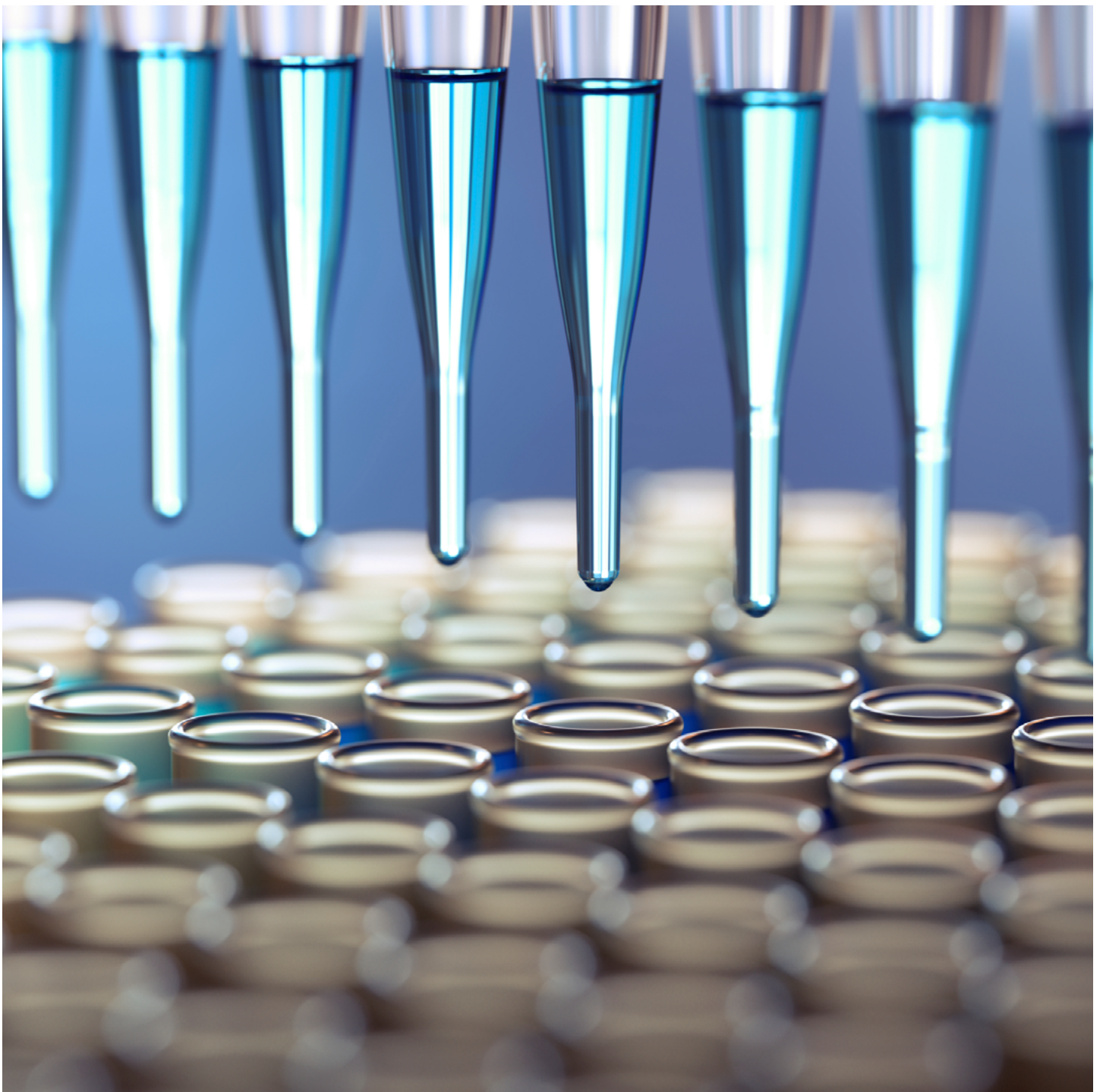


Public Health
England

Protecting and improving
the nation's health

Implementing pathogen genomics

A case study



Authors

Kathie Grant, Claire Jenkins, Cath Arnold, Jonathan Green, Maria Zambon

We acknowledge the contributions of Sobia Raza, Laura Blackburn, Louise Gaynor and Mark Kroese

Contents

	Background	4
1	Why develop pathogen genomics services for public health microbiology?	5
2	PHE Colindale WGS service	7
3	Facilitating implementation and delivery of the WGS service	12
4	Outcomes and impact of implementing the WGS service	22
5	Future developments	28
6	Conclusions	31

Background

Pathogen genomics is widely acknowledged as a powerful approach towards the investigation and management of infectious diseases. A number of initiatives and publications including the Chief Medical Officer's 2016 Annual Report¹ note the transformative potential of whole genome sequencing (WGS) to the delivery of microbiology and public health functions. At the same time these reports also emphasise the importance of changes in the design, operation and workforce of laboratories, in order to adopt this technology into routine practice.

Public Health England (PHE) has been working to implement WGS techniques since 2012. Its National Infection Service has a WGS strategy to enhance the control of communicable diseases nationally and to help PHE meet its regional, national and international obligations in infectious disease management. At PHE Colindale a centralised WGS service has been established that now provides a service for a range of pathogens of national importance including *E. coli*, *Shigella*, *Listeria*, *Campylobacter*, *S. aureus*, *Salmonella* and *Mycobacteria* (initiated by Birmingham). The WGS service at Colindale was one of the first UK laboratories to be offered accreditation for its whole genome sequencing services of bacteria and viruses.

Among the early adopters of the WGS approach for routine service delivery at PHE was the Gastrointestinal Bacteria Reference Unit (GBRU) – the national reference laboratory for gastrointestinal bacterial pathogens. This laboratory provides specialist testing of bacteria causing gastrointestinal (GI) disease, illnesses associated with foodborne transmission and those via direct contact with animals and their environment. The laboratory has now transitioned to using WGS as the routine typing tool for five major GI pathogens.

This document outlines the extensive development process undertaken by PHE to establish the central WGS service and the transformation of a national bacteriology reference laboratory into a genomics-led service. This includes changes in workflows, working practices and revisions to analytical procedures as services have migrated to using WGS for pathogen typing, surveillance and outbreak investigation. The application of genomics is reducing the number of people who become ill from foodborne infections.

¹ <https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2016-generation-genome>

1 Why develop pathogen genomics services for public health microbiology?

It is now well recognised that sequencing the whole genome of a pathogenic organism provides the highest possible resolution information which can be used to determine:

- Pathogen identity
- Relatedness to other organisms
- Sensitivity/resistance to drugs
- Virulence
- Identifying and tracking emerging or new infectious threats

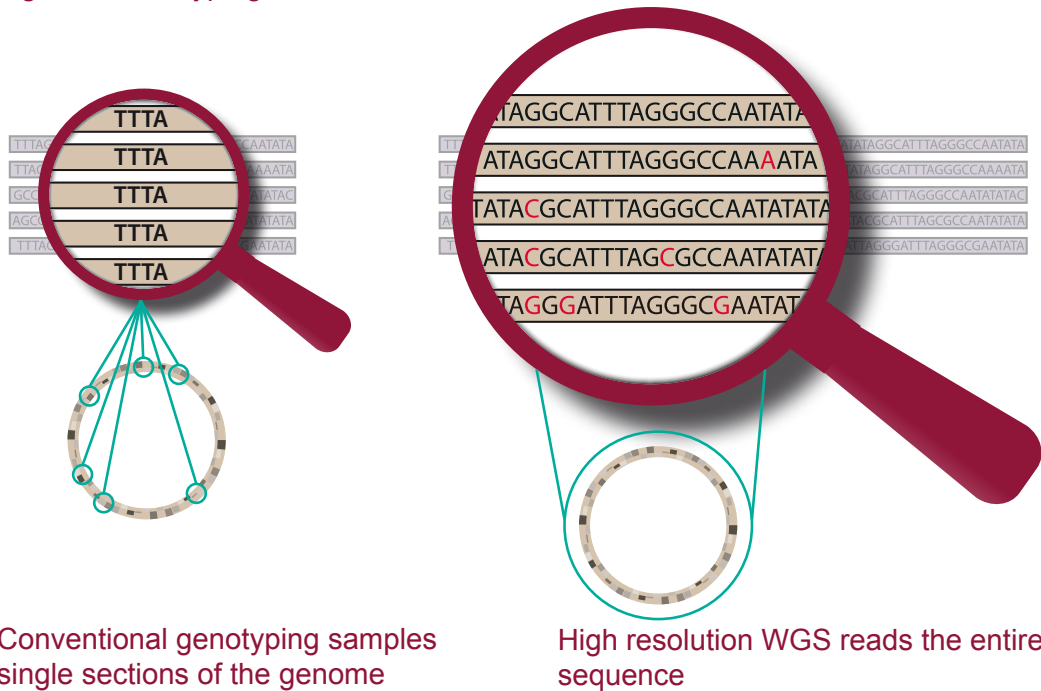
A major advantage of WGS is that this one technique can provide a more comprehensive set of microbiological information than multiple traditional laboratory processes. This minimises the need for laboratories to maintain the broad array of equipment, reagents and expertise to perform and analyse the numerous different phenotypic and genotypic assays needed to undertake surveillance of a diverse range of organisms.

Furthermore, WGS provides higher resolution information than conventional genotyping tests, which only analyse a limited portion of the genome, meaning that crucial information can be missed (Figure 1). Additional benefits include value for money, improved quality and laboratory health and safety and a reduction in animal use.

The transition to WGS use at PHE

The transition to the routine use of WGS at PHE occurred in response to the advantages that WGS provides over classical laboratory methods and the opportunity WGS provided to streamline laboratory methods whilst improving the management of infectious diseases.

Figure 1 Genotyping vs. WGS



The broader objectives and drivers for establishing the wider WGS service were:

- To provide a cost-effective, resilient and accredited WGS service for PHE
- To facilitate rapid adoption of WGS for microbiology services
- To ensure PHE is self-sufficient in WGS capacity for response to national incidents and emergencies
- To implement lean laboratory methodologies and reliable sample handling through robotics
- To better respond to outbreaks
- To stay at the forefront of the major shift in genome-based sequencing as a tool for clinical laboratory services
- To ensure international compatibility and exchange of information with public health institutes worldwide, many of whom are adopting WGS based investigation of infectious diseases

2 PHE Colindale WGS service

A centralised high throughput sequencing laboratory service has been established whereby extracted bacterial DNA is received, sequencing is performed and short read sequences delivered to the requesting reference laboratory for further analysis.

2.1 Which pathogens to sequence?

A review of existing microbiology reference services took place to identify the pathogen that would provide maximal benefits by transitioning to WGS, in terms of simplifying workflows, improving laboratory safety and reducing turnaround times. *Salmonella* was chosen to switch first because:

- The reference laboratory receives more *Salmonellae* than other gastrointestinal bacterial pathogens and transitioning to WGS presented the biggest improvements in terms of laboratory safety and simplifying workflows
- Traditional methods for characterising (sub-typing) *Salmonella* require a number of separate workflows that are lengthy and costly requiring extensive expertise and resources (Figure 2). There are complex multidirectional workflows taking place in different laboratories, with a large component of open bench work
- Each strain can be handled many times in order to carry out these tests, and the Health and Safety risk management of these processes made working practice cumbersome
- There was an opportunity to reduce the use of animals for producing the traditional test reagents

Once *Salmonella* WGS had been implemented, other enteric organisms could then be added to the WGS workflow.

2.2 Why are enteric bacterial samples sequenced?

Pathogen genomic data can inform and expedite the delivery of appropriate public health protection interventions by being used for:

- Pathogen surveillance – monitoring trends and the effectiveness of control measures
- Understanding how pathogens evolve
- Disease control, transmission analysis (who or what infected who), outbreak analysis (origin of outbreak: is an outbreak new, or related to an older outbreak?)
- Cataloguing antimicrobial resistance (AMR) with specific focus on AMR imported to the UK and AMR transmitted via food and/or contact with animals
- Tracking unknown or emerging threats such as foodborne and zoonotic GI disease

Data provided by the reference laboratory is used by infectious disease specialists including consultants in communicable disease who manage health protection at a local level.

WGS service users are:

- Health Protection Teams including consultants in communicable disease
- Local authorities including environmental health officers
- Consultant microbiologists in hospitals and Public Health microbiologists
- Food Standards Agency
- Animal and Plant Health Agency, DEFRA
- European Centre for Disease Control and European Food Safety Authority and international public health agencies

What is being sequenced?

More than 100,000 bacterial and viral genomes have been sequenced since the service launch in 2014. This is around 600-700 bacterial samples per week, on a routine basis, to a specified schedule.

Within the reference laboratory, WGS has been expanded to investigate all *E.coli*, *Shigella*, *Salmonella*, *Listeria* and *Campylobacter* samples (Table 1). The *Salmonella* service was the first to completely transfer to WGS – routine serotyping stopped in April 2015. Approximately 10,000 *Salmonella* isolates are received by the GBRU per year. WGS services for other pathogens of public health importance were launched shortly after the *Salmonella* service (Table 1).

Table 1: Dates of service switch to WGS, including numbers of pathogens sequenced by WGS since the central service launch in 2014.

Pathogen	Switch to WGS	2014	2015	2016	2017	Total
<i>Salmonella</i>	April 2015	7,600	10,100	10,000	9,400	37,100
<i>E. coli</i> and <i>Shigella</i> [§]	June 2015	-	3,000	4,300	4,500	11,800
<i>Listeria</i>	December 2015	-	-	1,000	1,400	2,400
<i>Campylobacter</i>	January 2016	40	450	350	450	1,290

[§]*E. coli* and *Shigella* figures are combined as they follow the same analysis pipeline

2.3 How does the WGS workflow operate?

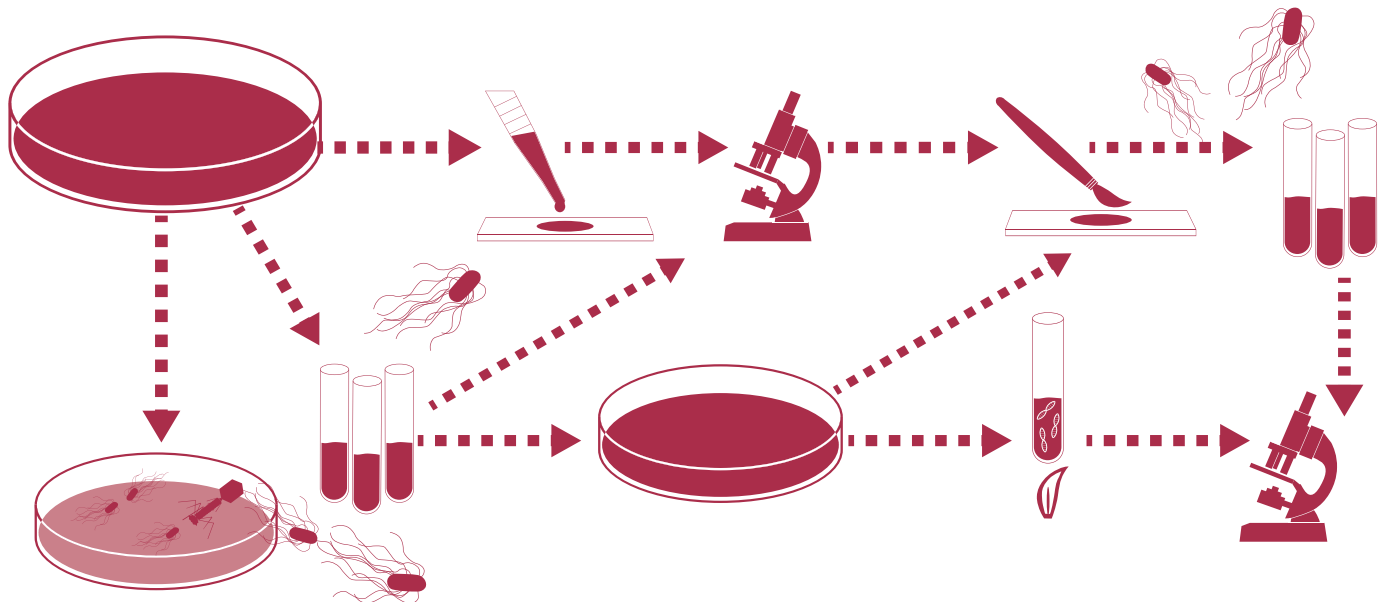
The principal working advantage of WGS is that the testing of all organisms can follow the same workflow, which greatly simplifies laboratory practices and facilitates improved safety management. Pathogens undergo minimal handling during WGS which reduces the risk of health and safety incidents. Once organisms have been grown in culture media, all sample processing is performed in a microbiological safety cabinet – an enclosed work area from which the air supply is extracted – in a category level 2 or 3 (biosafety level) laboratory. Pathogens are made safe by heat inactivation, which also helps release their DNA. Automated DNA extraction from the inactivated pathogens is performed in a second laboratory.

DNA samples are then sent to the central sequencing service where they travel through the NGS service workflow – robotics (to check quality and purity and for library preparation) and sequencers – while information about samples is collected via a laboratory information management system (LIMS). Sequence data are stored and analysed on the IT and bioinformatics infrastructure, then combined with LIMS data to produce the final report on a sample (Figure 2).

The WGS services takes five days from sample submission to final report.

Figure 2: *Salmonella* traditional methods vs. WGS

Conventional typing: before

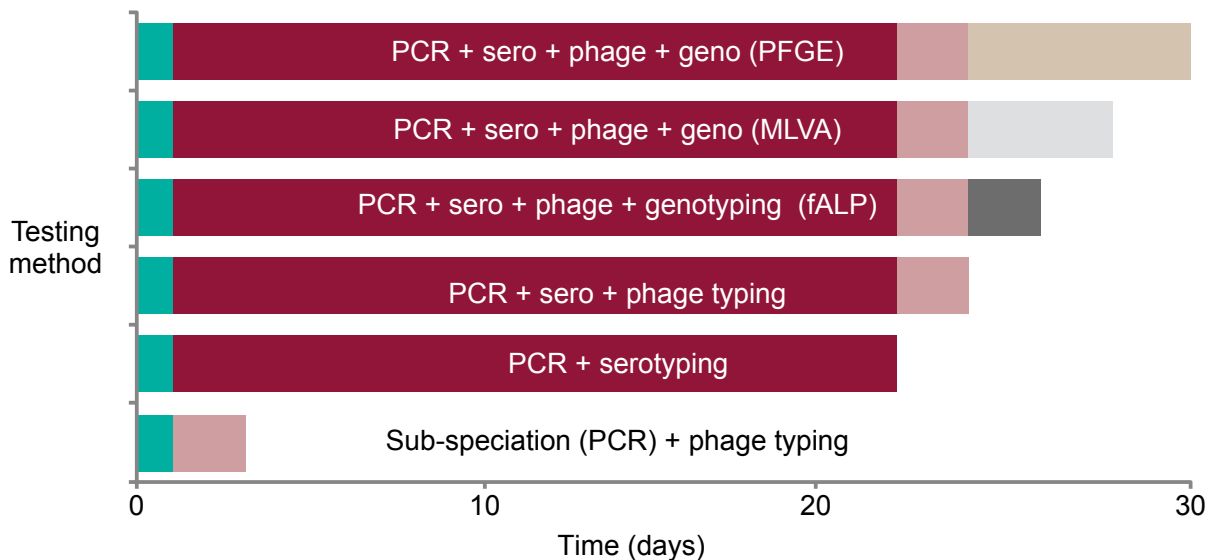


Once organisms have been grown in culture, all samples are processed within the GBRU laboratory utilising a range of testing methods.

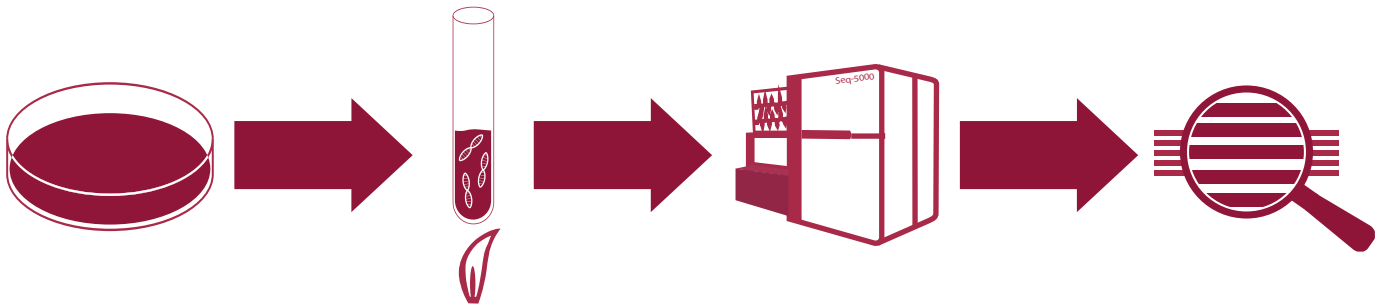
Methods include PCR, serotyping and phage typing followed by molecular typing methods such as PFGE, MLVA or fAFLP. The combination of these methods results in significantly longer turn-around times compared with WGS.

Serotyping of *Salmonella* is a long process that takes around 20 days and requires antibodies raised in animals. Genotyping methods that only examine restricted portions of the genome were reserved for outbreak investigations once available evidence strongly suggested an outbreak.

How long does conventional typing take?



Whole Genome Sequencing: now

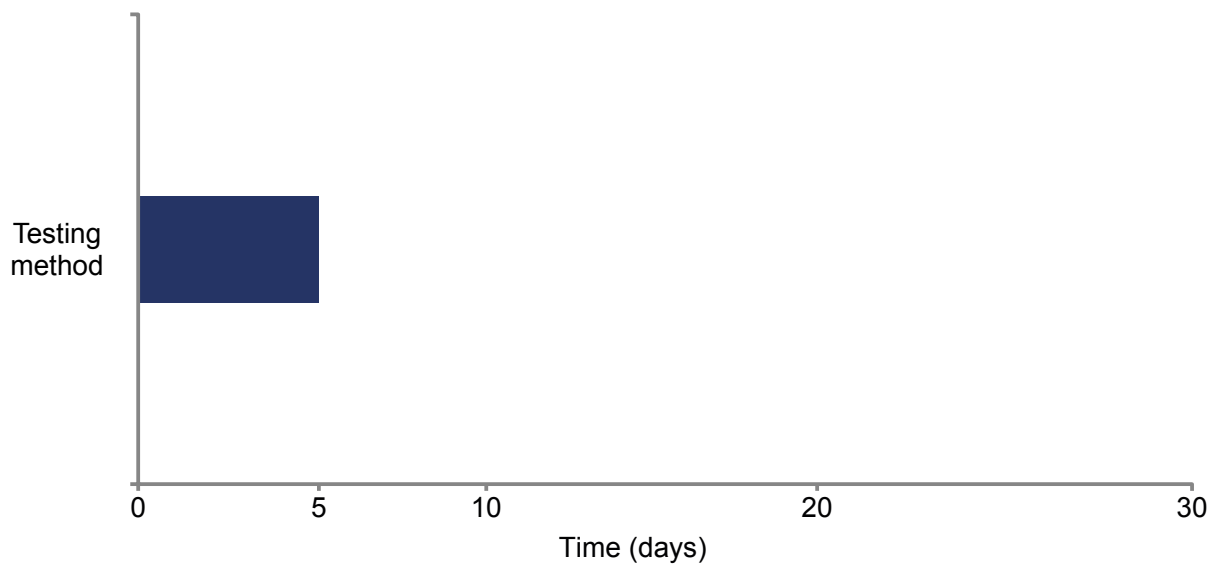


Once organisms have been grown in culture, all samples are processed to extract their nucleic acid (DNA).

Each nucleic acid sample follows the same streamlined workflow, generating a report containing organism identification and relatedness to other organisms within five working days.

WGS-based methods have vastly improved the identification and classification of *Salmonella* subspecies and serovars and identification of clusters during outbreaks. This enables better public health action to be taken to control and prevent further illness.

How long does the WGS service take?



3 Facilitating implementation and delivery of the WGS service

Implementing WGS is part of a broader PHE strategy to exploit genomics to improve infectious disease diagnosis and management and deliver better public health outcomes. The use of WGS represents a quantum shift in the technologies used to deliver reference laboratory services, replacing a variety of methods some of which have not changed since the 19th century. Delivering a WGS service of this scale is a significant undertaking, requiring clear organisational commitment for the investment, planning, collaborative working, and workflow revisions needed.

3.1 Planning WGS service requirements

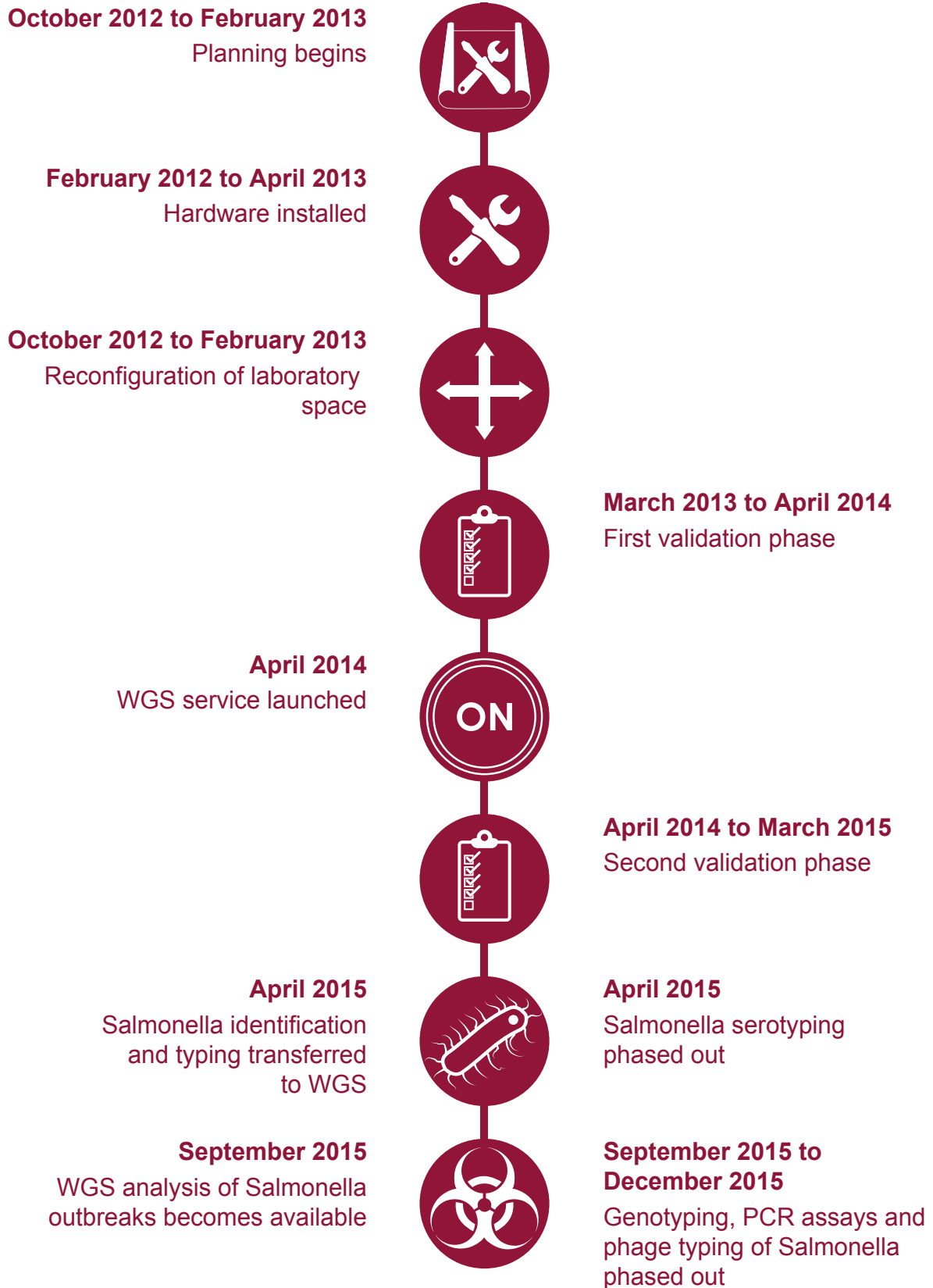
Business and capital planning

One of the first steps in the development process was to establish a robust business case for launching a WGS service with capital expenditure planning determining the costs of equipment and infrastructure (Table 2).

Table 2: Capital expenditure profile to support WGS implementation

Elements	Purpose	Cost (£, thousands)
Sequencing equipment and robotics	Capacity to provide high throughput 24/7 service with surge capacity and speed for emergency response	1,680
IT, data management and project management	<ul style="list-style-type: none"> • Sample collation and tracking • Data integration, processing and storage • Direction and oversight of implementation of the WGS service 	1,466

Figure 3: Implementation timeline for WGS service including *Salmonella*



Workflow planning

Given that WGS implementation would involve significant changes to existing methods and workflows the planning process then required:

- Cost analysis of traditional services vs new WGS services. Estimates of WGS service costs took into account costs of DNA extraction, sequencing, and costs associated with sample issues or sequencing failure
- Remodelling of laboratory space
- Review of health and safety requirements
- Arrangements for the procurement and installation of equipment and IT infrastructure (3.2)
- A review of workforce and training needs (3.5)

Establishing service requirements

Another key planning element was to understand the WGS service requirements and how service users would interact with it. This included determining:

- The processes needed for users to provide the appropriate quality and quantity of samples for the WGS service and their training needs
- The anticipated sample numbers, and turn-around-times required
- The process and format in which to return data to users, e.g. as raw-sequence data, or analysed by a bioinformatics pipeline
- How services might evolve, for example as the underpinning technologies develop

3.2 Reviewing and acquiring equipment

The preparation of samples for WGS required the development of a single, large-scale, automated workflow which replaced many small-scale and diverse manual operations. This shift in sample handling required major process redesign and specialised equipment to support the large-scale and high-throughput laboratory workflow and downstream handling of sequence data.

Planning equipment needs involved not only acquiring all necessary items but also building resilience into the system to ensure that enough equipment is available should one or more machines break down, and developing a business continuity plan to ensure that a service is available off-site if required.

Beyond the high-throughput and benchtop sequencing machines (which provide sequencing capacity for a 24/7 service and surge capacity), the other significant elements of infrastructure included:

Robotics

The ability to process a large number of samples in parallel is essential to providing a high-throughput centralised NGS sequencing service and for maximising outputs and reducing costs and turnaround times. Robotics systems facilitate workflow automation to minimise human error and improve processing consistency. They are used for library preparation, and to prepare standardised, quality controlled samples for sequencing. The plan for automated liquid handling included consideration of how robotics requirements may change in the future with developments in technology and introduction of new workflows, kits, or software.

IT infrastructure

A sequencing service requires extensive IT infrastructure to manage and store the data generated – this includes computer hardware, operating systems and software. Dedicated expertise were required to deliver the high-performance computing cluster and for the maintenance and migration of current infrastructure.

It was vital to ensure that the infrastructure meets current needs such as having enough data storage and a sufficient data processing speed that does not impact on service delivery, including being scalable to meet future needs.

3.3 Workflow development

In order to operationalise the WGS service at scale it was necessary to consider how different elements of infrastructure and the various workflow stages – from sample receipt to the reporting of results – should interact:

Specimen receipt and management including a Laboratory Information Management System (LIMS)

For the service to operate effectively, detailed guidelines were developed to direct users through sample preparation, robotics processing and sequencing, including a trouble-shooting guide. The interaction between the different parts of the system were considered: how samples are managed and tracked, including by the service users, and within the NGS service. The interaction between LIMS and the bioinformatics infrastructure is also important since it plays a key role in the delivery of the report to the users.

Bioinformatics

Central to the operation of an effective NGS service was the development of bioinformatics pipelines to analyse sequence data and return comprehensive results to users in a timely manner. This required changes in staffing and training, including recruiting new staff members to operate the service, since launch of the WGS service increased emphasis on, and workload related to, data science and management within PHE.

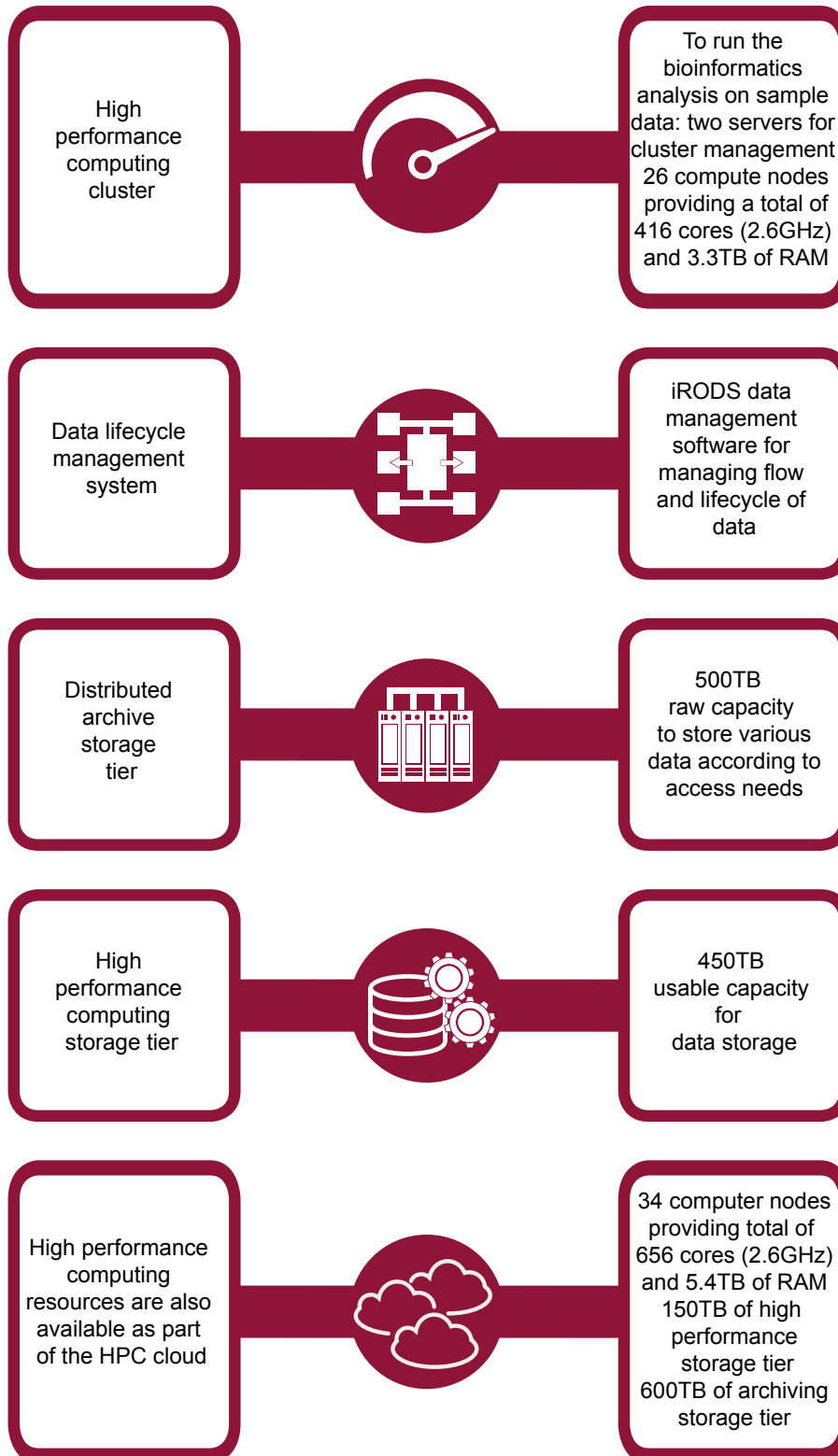
Development of analytical pipelines

In order to deliver routine WGS analysis for a wide range of microbial and viral pathogens, substantial development work was undertaken to understand the diversity within species using previously well-characterised, representative strain collections. It was essential to demonstrate correlation between outputs from bioinformatic analyses and pre-existing 'gold standard' typing methods and, where a lack of correlation was identified, the reasons fully investigated.

Development of the bioinformatics processes underpinning the WGS service was a significant undertaking, in order to provide an 'end to end' process including:

- Extraction of all sequence data from multiple samples from the sequencing machines – up to 96 samples can be sequenced at the same time – and categorisation of data about individual samples
- Reporting on the accuracy and quality of the data and integrate the ability to add this information once the sequence data has been collected
- Configuring files for data storage and providing service users with access to the data, including the processing required to give users the results that they need
- Deliver analysis pipelines, that can manage unexpected or unusual samples
- Considering data integration needs – how the sequence data can be integrated with other data about the sample (e.g. when and where it was collected, type of sample) which gives epidemiological context
- Designing reports to return information to users and communicating to users how to understand reports.

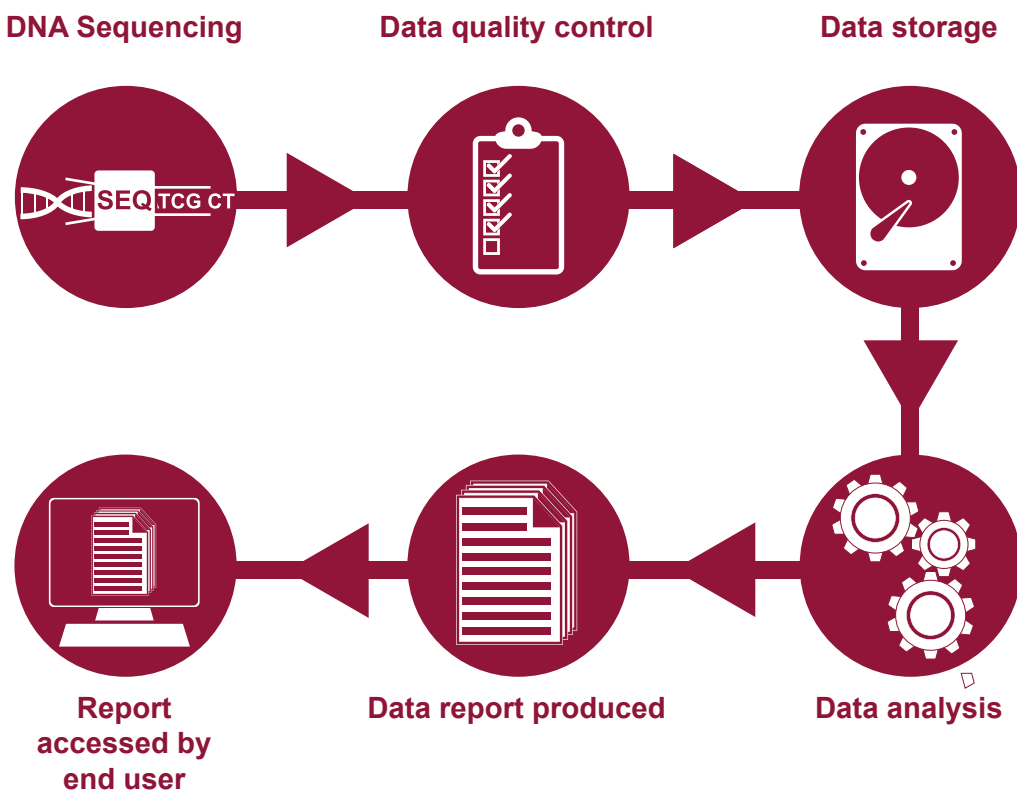
Figure 4: Informatics infrastructure requirements



Delivery of these components was supported by:

- A Laboratory Information Management system built on open source software to carry out many of the functions required
- A compute cluster that can deliver computationally intensive data processing
- High performance storage to meet the needs of expanding use of WGS

Figure 5: Bioinformatics data pipeline



3.4 Validation and trouble-shooting of all processes

Even with the equipment and workflows in place, before traditional testing could be switched off it was essential to validate the WGS service to ensure that the process was reliable and reproducible. In practice this meant:

- Parallel running of both conventional methods and WGS workflows and comparison of datasets to verify performance characteristics
- Undertaking gap analysis and adjusting protocols accordingly
- Establishing quality parameters for each step of the entire WGS-based typing process
- Assessing the quality of the input and output material (i.e. nucleic acid and analysed genetic information) to determine whether both were providing high-enough quality information, performing as expected, and matching or exceeding performance of existing methods

Phase I of the validation compared phenotyping and WGS of 1538 representative *Salmonella* samples from 2012 to establish if WGS could be reliably compared to phenotyping. Phase II (April 2014-15) involved comparing the phenotypic serotype of 6887 *Salmonella* isolates to the WGS-determined serotype.



97% concordance between WGS and conventional methods



Novel genotypes identified that needed phenotype characterisation

During the validation process the types of issues identified explaining the 3% discordance and managed through adjustment of operating protocols included:



Not all known *Salmonella* serotypes were assigned a MLST profile



Different serovars belonging to the same sequence type



Processing errors in laboratory phenotyping serotyping



Genetic sequences incorrectly predicting the phenotype



Optimisation of DNA extraction procedures for quality and quantity

3.5 Staff training and engagement

The switch to WGS required an increase in multi-disciplinary working, changes in working practice, and therefore in training requirements for staff working in laboratories and for those analysing and managing data:

- Laboratory staff required training to prepare samples, operate the equipment that analyses the samples, gain knowledge of sequencing and some understanding of the bioinformatics methods.
- Specialised bioinformatics staff were needed to develop bioinformatics pipelines and analyse data, as well as staff with expertise in computing and software development.
- Expert bioinformaticians in PHE are passing on skills and expertise to other users such as epidemiologists and clinical scientists who wish to undertake genomic data analysis as part of their roles.
- Establishment of a Galaxy system for laboratory staff to explore use of bioinformatics software and to support research activities
- An online learning resource, [ePathGen²](http://public-health-genomics.phe.org.uk/), is available

Engagement efforts with staff and service users to inform them of service changes and explain how WGS information would be returned took place via:

- Information letters and surveys
- Roadshow events run in different regions of the country by reference laboratory bioinformaticians, and other senior laboratory and epidemiology staff for front line public health staff who would be receiving WGS data on *Salmonella* and other pathogens. Information included the basics of sequencing, how WGS is used e.g. outbreak investigation and details on results databases
- Collaborations with other agencies such as the APHA and FSA and presentations at external meetings contribute to raising awareness of pathogen WGS (Figure 6)
- International public health collaborations on WGS implementation, with ECDC and CDC

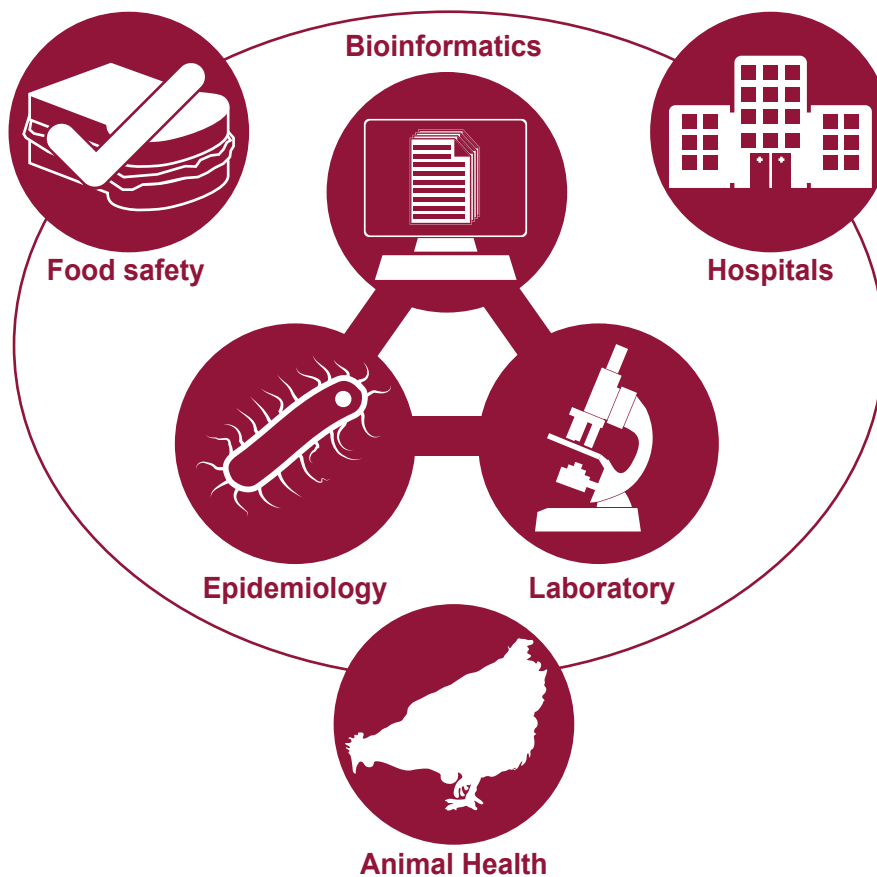
2. <http://public-health-genomics.phe.org.uk/>

3.6 Accreditation

The central sequencing service is now accredited to ISO15189 for WGS (UKAS lab reference number 8727) and the reference laboratory for services for *Salmonella*, *E.coli* STEC, *Campylobacter* and *Listeria* using WGS (UKAS lab reference 8197).

Whilst public health reference laboratories have a significant history of accreditation, this is a new challenge for bioinformatics as the traditional metrics used for microbiology are not as easily applicable to bioinformatic pipelines. However, the emergence of diagnostic genome sequencing based approaches for human disease in clinical laboratories has already brought bioinformatics into the accreditation scope and, as such, act as a valuable template for pathogen genomics. The applications of standards for WGS for public health microbiology has been a challenge for both PHE bioinformaticians and for UKAS but both have learnt much in recent months and a clear picture of requirements is emerging.

Figure 6: WGS facilitates multidisciplinary working and cross-agency collaboration



4 Outcomes and impact of implementing the WGS service

4.1 What has been the operational and other collateral impact of the WGS service?

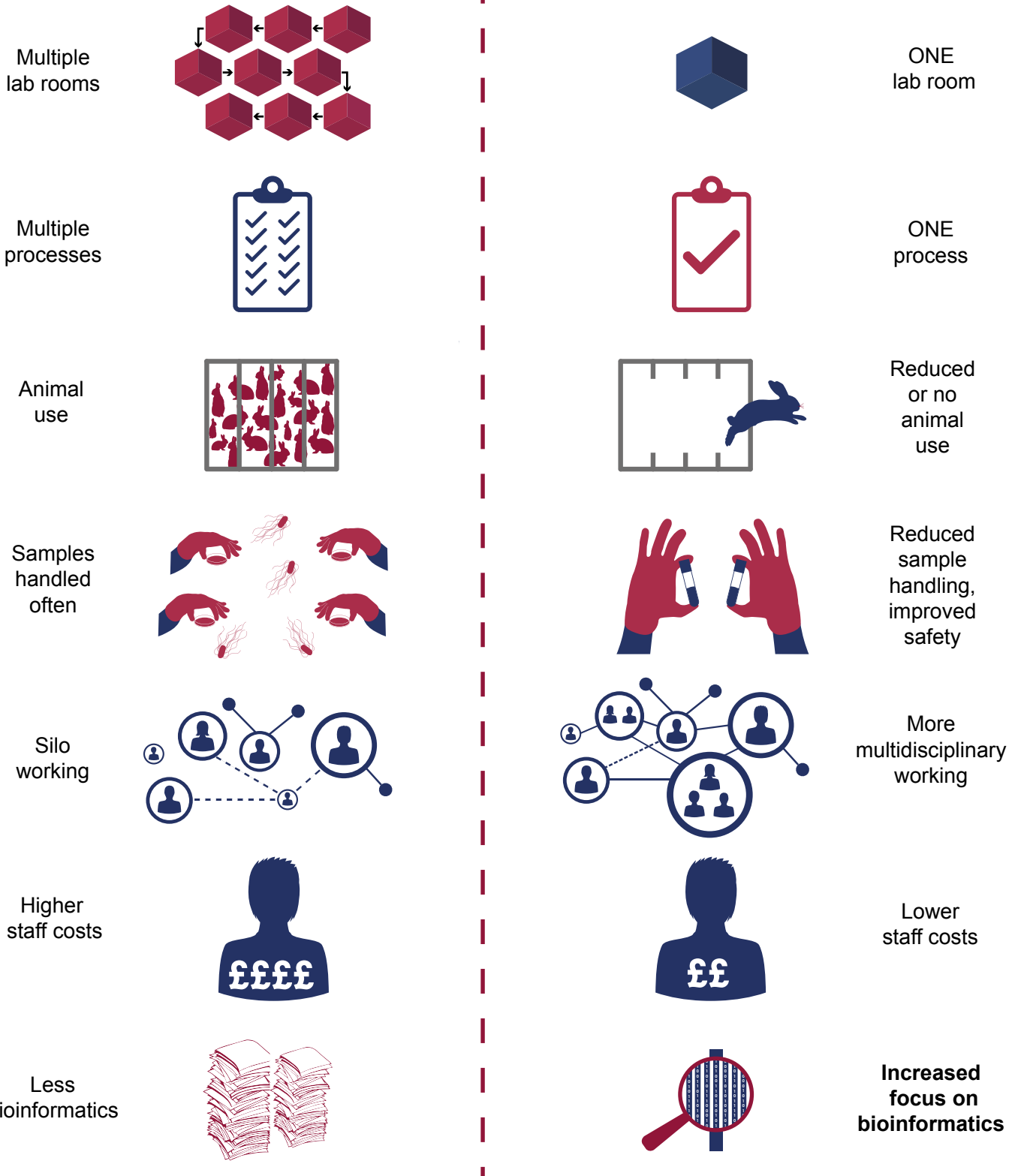
The significant benefits and impacts of the WGS service have been reduced costs and improved workflow processes and safety, including:

- More streamlined laboratory processes:
 - WGS has replaced at least 10 different validated processes for different bacteria
 - Pathogens are now typically processed in fewer rooms (e.g. *Salmonella* now travels only through one rather than nine different laboratories)
 - Whilst laboratory incidents were rare, the risk has been further reduced. Before WGS, most samples containing live organisms would be handled between 7 to 9 times, each step potentially exposing a laboratory worker to a pathogen. After implementing WGS, all samples require handling twice, and fewer than 5% require handling for another reason, e.g. WGS failure or ambiguous results
 - Reduction in staff costs of 17%
 - Reallocation of scientific staff from laboratory processing to data analysis and interpretation
- Reduced need for animal use – e.g. number of rabbits used for *E. Coli* and *Salmonella* work has dropped from 20-30 / year to zero
- An increased focus on data and data science due to the bioinformatics analysis of sequencing data
- Greater requirement for and emphasis on multi-disciplinary working between scientific colleagues as well as clinical and epidemiological staff
- WGS data gives improved microbiological understanding highlighting areas where more information could be collected for tracking the source and transmission of infection

Figure 7: Impact of WGS on laboratory service

Before WGS

After WGS



4.2 Genomics in action

In terms of public health value, the impact of the WGS service has been:

- Resolution of outbreaks that were not being solved using conventional methods
- Quicker resolution of outbreaks ensuring rapid interventions preventing ongoing transmission
- Greater resolution of information including detection of cryptic³ outbreaks not detectable using conventional methods
- Highly discriminatory, forensic level typing identifies robust links between cases and the source of the outbreak
- Evolutionary context provided by WGS data analyses reveals common sources of infection associated with seemingly sporadic cases
- Improved understanding of the transmission pathway of enteric pathogens

The principal benefits for the reference laboratory have been:

- Improved laboratory management
- Simplified staff training
- Improved workflows
- Improved health and safety
- Increased knowledge and expertise in pathogen genomics

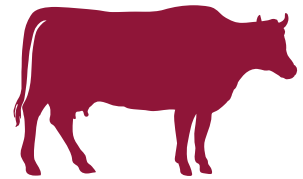
3. A cryptic outbreak is an outbreak where a low number of cases occur in time and space making them difficult to detect. Transmission in this type of outbreak can be from unexpected sources (e.g. environmental reservoirs of infection) and challenging to track

Genomics in action 1

Outbreak of *Escherichia coli* from raw (unpasteurised) drinking milk

Threat

Shiga toxin-producing *E.coli* (STEC) O157 are a major cause of food-borne illness and can cause severe symptoms. One outbreak of STEC O157, caused by consumption of raw drinking milk (RDM), was linked to a farm in South West England.



Action

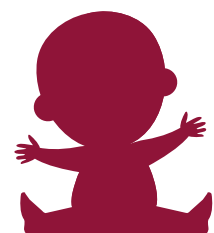
WGS analysis showed that the outbreak strain was present in faecal specimens from dairy cattle on the farm, and also identified additional cases, who initially did not report drinking RDM, as part of the outbreak.

Despite the initial failure of the cases to recall their consumption of RDM, high-resolution WGS data helped the investigators to identify these geographically dispersed cases. Subsequent further questioning of these cases revealed that they had, in fact, consumed RDM from the implicated farm.



Outcome

Action could be taken quickly: distribution of RDM from the farm was stopped, and public health messages were issued by the Food Standards Agency highlighting the potential dangers of consuming unpasteurised dairy products, particularly for children and people with underlying health conditions.



See <https://doi.org/10.1017/S0950268816000509>

Genomics in action 2

International outbreak of *Salmonella* Enteritidis from contaminated eggs



Threat

Near real-time WGS was used to assess the origin of an outbreak of *Salmonella* Enteritidis in a hospital in central England, affecting 32 inpatients and contributing to the death of one individual. Determination of the strain characteristics by WGS enabled the outbreak control team to promptly associate this outbreak with other cases of *Salmonella* Enteritidis linked to restaurants in other parts of England, and to outbreaks in France, Germany, Luxembourg and Austria.



Action

Including sequencing data in epidemiological and environmental investigations for the first time during an ongoing outbreak enabled the team to trace its origin to a German egg producer. This process would have been difficult, if not impossible, using conventional methods.



Outcome

The findings informed public health control measures to ensure the appropriate processing of contaminated eggs in Germany, and prompted food hygiene guidelines to be reissued to affected establishments to minimise the impact of cross-contamination.

See <https://doi.org/10.2807/1560-7917.ES2015.20.16.21098>

Genomics in action 3

Outbreak of *Salmonella* Enteritidis from mice used to feed reptiles

Threat

Salmonella Enteritidis causes most cases of salmonellosis (diarrhoea, fever, cramps and vomiting) in the UK, and is commonly transmitted by contaminated poultry and eggs. One of the challenges of investigating outbreaks using conventional methods is that low level, persistent outbreaks, on-going for months or even years but with a relatively small number of cases each week, tend to be masked and not detected. These outbreaks occur 'under the surveillance radar', meaning that specific outbreak investigations are not initiated.

Action

Due to the implementation of WGS such outbreaks are now being detected. One example is a 'slow-burn', low-level outbreak of *Salmonella* Enteritidis that had been ongoing for four years with 40% of cases affecting children under 10 years old. PHE and colleagues at the European Centre for Disease Control and the Animal and Plant Health Agency worked together using the high-resolution information provided by WGS to trace the outbreak back to *Salmonella*-infected mice being sold and used as feed for pet snakes.

Outcome

As snakes are not classified as pets, feeder mice are not considered pet food and are not subject to the same regulations regarding pathogen screening of pet food – the team therefore recommended that screening of feeder mice should take place to ensure they are free of *Salmonella*.

See <https://doi.org/10.1016/j.fm.2017.04.005>



5 Future developments

5.1 What next for WGS services?

The use of better established whole genome sequencing technologies and of newer portable long-read nanopore sequencers, combined with epidemiological information, has provided greater insight into the origin and spread of disease outbreaks that would not have been possible using standard techniques. The key to strengthening and optimising services in the future will be:

- Studying the effect that WGS has had on reducing gastrointestinal illness and the impact of genomics on detecting and resolving disease outbreaks
- Determining the cost-effectiveness of sequencing different pathogens
- Ensuring that our current analytical approaches (and underlying compute capability) are scalable to meet the increasing number of strains to be accumulated in the coming years through the development of robust scientific enterprise software
- Developing strategies for global capturing and sharing of genomic and related surveillance data
- Developing improved approaches for the visualisation of cluster and other data to the end-users
- Exploiting the sequencing data being generated
- Exploring the potential impact of new sequencing technologies

This will be facilitated by the sharing of pathogen sequence data – all of the *Salmonella* sequences generated by the reference laboratory are uploaded to the international public archival resource the 'Sequence Read Archive (SRA)' through the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/sra).

The power of portable long-read sequencing

New long-read sequencing technologies such as the MinION nanopore sequencer, are expected to impact on infectious disease management because of:

1. Portability: The sequencer is the size of small mobile phone and has been used in the field
2. Speed: sequencing is very rapid using this technology enabling sequence analysis to begin as soon as sufficient data has been generated and before the full sequence has been determined
3. Utility: Long read sequencing overcomes some of the challenges of using short-read sequencing to analyse bacterial genomes

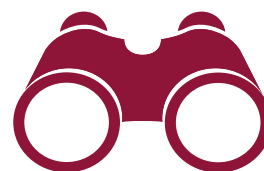
Laboratories can now determine the presence and location of bacterial virulence and antibiotic resistance genes, with precision, within clinically relevant time-frames. This information is important for determining how resistance spreads through bacterial populations, how highly pathogenic strains emerge, for rapid detection of outbreak strains and the impact this has on public health.

How does travel affect the gut microbiota?

A study is underway using WGS to study the presence of antimicrobial resistance (AMR) in the gut microbiota of recent travellers to countries where the prevalence of AMR is high and will also look at the impact of antibiotic treatment on the persistence of AMR elements in these bacteria. Studies such as this will increase understanding of the spread of antimicrobial resistance in an increasingly inter-connected world.

Which drug for which bug?

Research has shown that antimicrobial sensitivity information for *Shigella*, *E. coli* and *Salmonella* predicted using phenotypic methods and using WGS have a very high level of agreement. Ongoing evidence collection including a reference database of antimicrobial resistance genes will provide further insight into the genetics of drug resistance and enable future uses of WGS for antimicrobial susceptibility testing.



See <https://doi.org/10.1093/jac/dkx170>, <https://doi.org/10.1093/jac/dkw371> and <https://doi.org/10.1093/jac/dkw093>

5.2 Understanding what future service configuration might look like

With the advent of WGS services and developments in portable long-read genome sequencers, another area for consideration is the impact of technological developments on the delivery, configuration, and use of services in the future. With a growing range of technological capabilities, key questions raised include:

- Where and by whom is sequencing done – laboratory or in the field?
- Data – how are data collected and shared? Who has oversight? Who has access to the data and for what purpose?
- When and how will the available sequencing platforms be used – for outbreak management in the field in addition to current surveillance use?

6 Conclusions

The successful and effective implementation of the central sequencing service by PHE has had a significant impact on the work of the GBRU reference laboratory and other units within PHE.

The high degree of automation, consistency and safety ensures that the nucleic acid extracted from each pathogen is analysed via a streamlined process that minimises the number of procedures needed to determine pathogen identity and support outbreak investigation and surveillance activities. This has resulted in reducing consumable and equipment costs for multiple laboratory workflows, reducing the laboratory footprint and staffing costs. Implementation of WGS has refined outbreak investigations by providing more robust case definitions enabling cases to be ruled in or out of outbreaks more accurately.

The introduction of routine WGS has improved national and local surveillance, increased the number of outbreaks being detected and has led to outbreaks being detected earlier than previously possible. WGS analysis provides information on the potential source and/or potential geographical location of the source as well as improving monitoring of the effectiveness of control and preventative measures. Overall, the use of WGS in this way means that fewer people will become ill.

Future developments in sequencing technologies are likely to support more widespread use in all areas of medicine and public health.

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. We do this through world-leading science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health and Social Care, and a distinct delivery organisation with operational autonomy. We provide government, local government, the NHS, Parliament, industry and the public with evidence-based professional, scientific and delivery expertise and support.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: @PHE_uk
Facebook: www.facebook.com/PublicHealthEngland



© Crown copyright 2018

You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit OGL or email psi@nationalarchives.gsi.gov.uk. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Published August 2018
PHE publications gateway number: 2018254



Public Health England supports the
UN Sustainable Development Goals

