Institution: University of Birmingham

Unit of Assessment: 5 – Biological Sciences

Title of case study: Real-time pathogen genome sequencing to inform outbreak response

Period when the underpinning research was undertaken: 2013–present

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Prof. Nicholas Loman	Professor of Microbiology and Bioinformatics	1 April 2007–present
Dr Josh Quick	UKRI Future Leaders Fellow	2013 –present
Prof. Mark Pallen	Professor of Microbial Genomics	July 2001–March 2013

Period when the claimed impact occurred: 1 January 2009–31 December 2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

Loman's (UoB) development of rapid whole genomic sequencing methods, including portable sequencing tools, laboratory methods and open-source bioinformatic software, have transformed the management of **infectious disease in both developed and developing countries**, improving identification of transmission pathways, evolution of pathogens and sites of persistence. Used in the 2014–2016 Ebola epidemic in West Africa, genome sequencing **shortened the length of the epidemic reducing mortality, morbidity, and economic loss**. The approach was then applied to other diseases and notably the SARS-CoV-2 pandemic. The benefits led to **changes in policy and practice within the World Health Organization** to **improve rapid sharing** of sequence data which has proved critical in the identification of SARS-CoV-2 variants and **influenced containment policies** in the UK.

2. Underpinning research

Whole genome sequence data for epidemiological investigations

From 2008 to 2014, UoB researchers were the first to recognise and describe the potential for **using whole-genome sequence (WGS) data during outbreak investigations**. In 2008, the group investigated a hospital outbreak of a severe multiple-drug resistant strain (Acinetobacter) and determined that cases were imported by military personnel returning from Iraq and Afghanistan and transmitted to NHS patients in intensive care units [R1]. That was followed by the first large-scale demonstration of whole-genome sequencing to investigate a large outbreak of *Escherichia coli* in Europe which led to an intensive crowd-sourced effort to analyse the genomic sequence to determine its source: the group's investigations determined that the *E. coli* was most likely a human-circulating strain rather than the more usual zoonotic source [R2]. In both cases, control of the outbreaks was improved by knowing the transmission pathways.

Technological development in support of real-time outbreak detection

Due to equipment limitations, early investigations using WGS were done mainly in well-equipped labs in a handful of genome centres worldwide. However, by 2014, the Oxford Nanopore Technologies MinION sequence platform had become available which enabled rapid analysis of sequence data in the field. The 'pocket sequencer' weighs <100 grams, costs $1000 and can be powered through the USB port of a standard laptop. UoB researchers became the first to generate and publish data of any type from this technology (June 2014) and collaboratively worked with Oxford Nanopore Technologies (a private UK company) to develop laboratory methods and software for the platform as well as testing and providing feedback on products.

The group tested the device on a large hospital outbreak of *Salmonella* in Birmingham as an early example of **real-time genomic epidemiology** [R3] and developed prototype bioinformatics and laboratory methods to demonstrate how the MinION and software might be used in outbreaks of infectious disease. The study highlighted the need to develop bespoke bioinformatics to analyse the long, error-prone sequences generated by the MinION platform. To address this need, Loman developed an open-source software tool, Poretools, to process raw nanopore data [R4]. Loman then worked with Jared Simpson (Ontario Institute of Cancer Research) to develop the nanopolish approach to assemble a genome from nanopore MinION data, providing the **first demonstration of a complete *de novo* assembly of a microbial genome** (*Escherichia coli* K-12 MG1655) [R5].

Real-time outbreak detection and analysis in the field

Loman rapidly recognised that portable nanopore sequencing could have profound benefits for management of infectious diseases globally, resulting in a 'lab-in-a-suitcase' suitable for 'pop-up' WGS at the site of disease outbreaks, especially in countries where storing and shipping samples for analysis was not possible. To meet this aim, the group developed the first complete protocol for sequencing RNA viruses (Ebola and Zika) directly from clinical samples using nanopore technology [R6, R7]. This work included bioinformatics analysis tools which **enabled very accurate sequences to be obtained from this previously noisy technology for the first time**. Deployment of 'pop-up' sequencing laboratories (Figure 1) was first used in the 2014–2016 Ebola outbreak in Guinea, delivering genome sequencing data in as little as 18 hours to enable field epidemiologists to identify and break transmission chains.

Figure 1. 'Lab-in-a-suitcase' deployed to Conakry, Guinea in 2015 by Dr Joshua Quick (left). Portable genome sequencing facility in Nongo, Guinea in 2016 (right).

Broader applications of the real-time WGS approach

Following work on Ebola and Zika, Loman and team have continued to develop technologies, models and processes for epidemiological investigation of outbreaks through the realisation of **real-time, pathogen genome sequencing and rapid open data sharing**. The sequencing approach they invented, Primal Scheme and PrimalSeq [R7], has been used to analyse many other infectious diseases (including Yellow Fever, dengue, rabies, polio, influenza, West Nile Virus and chikungunya). Crucially, the underpinning approaches have been used by Loman and others to provide cost-effective means for tracking the evolution of SARS-CoV-2 [R8].

Key Findings:

KF1: Whole genome sequencing can provide a powerful tool to investigate outbreaks of disease (R1, R2, R5, R8).

KF2: The use of real-time nanopore sequencing is transformative in providing timely data for decision making to control infectious disease transmission but requires bespoke bioinformatics software to improve the accuracy of real-time nanopore sequencing (R3, R4, R7).

KF3: The combination of portable hardware plus open-source software for real-time analysis of viruses in patient samples in disease outbreaks can inform disease control decisions (R6, R7).

KF4: Real-time genome sequencing accompanied with rapid data sharing has become a standard approach for monitoring the transmission of infectious disease across both the developed and developing world (R6, R7, R8).

3. References to the research

R1: Lewis T, Loman NJ, Bingle L, *et al.* High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak. *Journal of Hospital Infection*. 2010;**75**(1):37-41. doi: 10.1016/j.jhin.2010.01.012

R2: Rohde H., Qin J., Cui Y., *et al.* Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. *New England Journal of Medicine*. 2011;**365**(8):718-724. doi: 10.1056/NEJMoa1107643

R3: Quick, J., Ashton, P., Calus, S. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella* . *Genome Biology* 2015;**16:**114 doi: 10.1186/s13059-015-0677-2

R4: Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics*. 2014;**30**(23):3399-3401. doi: 10.1093/bioinformatics/btu555

R5: Loman, N., Quick, J. & Simpson, J. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature Methods* 2015;**12:** 733–735 doi: 10.1038/nmeth.3444

R6: Quick, J., Loman, N., Duraffour, S. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;**530:** 228–232. doi: 10.1038/nature16996

R7: Quick, J., Grubaugh, N., Pullan, S. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols* 2017:**12:** 1261–1276. doi: 10.1038/nprot.2017.066

R8: Tyson, J.R., James, P., Stoddart, D., *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *BioRXiv* 4 Sep 2020 doi: 10.1101/2020.09.04.283077

4. Details of the impact

Impact on professional practice via the management of outbreaks

The use of real-time genomic epidemiology, based on KF1–KF4, has transformed the way that outbreaks of **infectious disease are controlled in developed and developing countries**. The approach enables epidemiologists to accurately track transmission routes in real-time and then intercept the chain to prevent further transmission of the virus. As such, **professional practice has been influenced by the research** with genome sequencing being used routinely during large outbreaks of national or international concern that constitute public health emergencies. This case study provides an example from both developing and developed countries.

1. Example from developing countries: Ebola and West Africa

The **control of the 2014–2016 West African Ebola epidemic was transformed** by the use of real-time whole genome sequencing [KF2, KF3], which was well suited to the epidemic of Ebola which demanded rapid case detection followed by contact tracing. The approach facilitated "rapid containment of the flare-up[s]" ultimately shortening the epidemic therefore bringing both **public health and economic benefits** to the region [EV1].

By sharing the genome sequence data produced directly with outbreak responders (KF4), the **policy and practice of the World Health Organization (WHO) was changed** [EV2]. The data informed epidemiological studies which were used to infer the source of new chains of transmission of Ebola in newly affected or previously 'cured' regions [EV3]. Fundamentally, the data helped to target outbreak response resource, for example surveillance and emergency response measures [EV1]. As testified by a technical officer of the WHO:

> The team were able to rapidly establish in-country sequencing capacity […] genomic data generated within 72hrs was used to rapidly identify survivor associated flare-up, expand the understanding of transmission chains and used to exclude new introductions of the virus during the outbreak. [EV2]

There were particular challenges in West Africa due to the geographical spread of the disease which had led to uncertainty on the origin of new cases. However, genome sequencing was able

to distinguish between sporadic cases and established chains of transmission [EV3]. Examples are provided in the WHO weekly situation reports:

> The case from Conakry […] is not a known contact of a previous case, and genomic analyses suggest he was not infected with the strain of Ebola virus responsible for the most recent cases in Conakry and Forecariah. [EV3]
>
> The first case identified from Forecariah, […] genomic analyses suggest she is part of the same chain of transmission — the Ratoma chain — as the 4 cases that were reported from the same subprefecture in Forecariah during the week. [EV3]

Sequencing also contributed to detection of novel routes of transmission and viral persistence (e.g. the potential for breastmilk to transmit Ebola from mother to child). However, this is best evidenced by the detection of active virus in a survivor's seminal fluid (over 500 days after the initial infection) which caused a new cluster of cases in Guinea, more than a year after the previous last case [EV4]. Such findings focused the epidemiological response and are **now included in WHO guidance** for clinical care for survivors of Ebola virus disease:

> Recent data suggest that Ebola virus can persist in the semen of males for a year or more after acute infection […] Consequently, all EVD survivors and their sexual partners should receive counselling to ensure safe sex practices until their semen has been determined to be free of Ebola virus. [EV4]

The WHO now recognise that "Pathogen genetic sequencing data is an increasingly valuable source of information to understand and control outbreaks of infectious disease" [EV5] and has used the approach [KF2, KF3] to tackle other serious outbreaks of priority diseases across the world, including the Zika public health emergency in South America during 2016, a severe Yellow Fever outbreak in Brazil in 2017–2018 and to determine the origin of Lassa fever cases in Nigeria in 2017 [EV5]. Furthermore, the methods and protocols developed have been begun to be used more widely for diverse infectious diseases including rabies, polio, African Swine Fever, TB, malaria, responsible for >2m deaths/year globally [e.g., EV6].

2. Example from a developed country: SARS-CoV2 and the UK: Identification of variants

KF4 has been utilised to analyse the evolution of strains of SARS-CoV2 from its first detection in late 2019 [R8]. Whilst early evolution of the virus prompted little alarm, in **December 2020 a new strain of SARS-CoV2 was identified which was of concern to parliamentarians and influenced policy** to contain the new threat.

The novel SARS-CoV2 'Kent' variant was identified in Dec 2020, using the approaches in KF3 and KF4, and was presented in a series of technical briefing documents [EV7]. These reports led to Parliament being briefed by the Health Secretary on the 14 December 2020 [EV8i] with the subsequent press conference led by the Prime Minister to outline a **reversal in planned policy** to relax restrictions on travel and group meetings planned for the 5 days around Christmas 2020 [EV8ii]. The data then underpinned planning conducted in December 2020 for the 3rd national lockdown in the UK in 2021:

> Over the past few days, thanks to our world-class genomic capability in the UK, we have identified a new variant of coronavirus, which may be associated with the faster spread in the south-east of England. Initial analysis suggests that this variant is growing faster than the existing variants. […] We have notified the World Health Organisation about this new variant […]. [EV8i]
>
> The UK has by far the best genomic sequencing ability in the world, which means we're better able to identify new strains like this than any other country […]. Given the early evidence we have on this new variant of the virus, the potential risk it poses, it is with a very heavy heart, I must tell you, we cannot continue with Christmas as planned. [EV8ii]

The contribution of Loman in helping to contain the spread of the virus was praised by the executive director of COG-UK (COVID-19 Genomics UK) in a letter addressed to the Vice Chancellor of the University of Birmingham:

> I am writing to acknowledge and extend my heartfelt gratitude for the support that the University of Birmingham has provided to the […] COG-UK consortium since its inception, a year ago this week. The work undertaken by your colleagues, including Professor Nick Loman and his team, has been vital to the success of the consortium and our contributions to the pandemic response in the UK and beyond. […] The recent emergence of SARS-CoV-2 variants has brought the work of all individuals and organisations involved in COG-UK to wider attention. This work will likely remain vital for some time to come as we continue to monitor for variants as the vaccination campaign proceeds [EV9]

Real-time WGS has changed policy and practice of the World Health Organization:

Following the proven use of genome sequencing in West Africa, the **practices of the World Health Organization were changed** by the adoption of an open-data policy. From 2016, the WHO has recommended that data should be shared within 21 days of sequence generation. This underlined the importance and relevance of genomic datasets to outbreak response. The WHO code of conduct now states:

> In each new outbreak or new transmission season genetic sequence data […] should be made publicly available. In particular the first set of sequences providing crucial information on the pathogen, genotype, lineage, and strain(s) causing the outbreak which may inform on the origin […] as well as the choice of diagnostics, therapeutics, and vaccines, should be generated and shared as rapidly as possible. [EV5]

The revised code of conduct has made the sharing of data a crucial tool in responding to new variants of SARS-CoV2 with over 500,000 genomes shared by 31 December 2020:

> Sharing of full genome sequences is facilitating detailed analyses by partners […]. Genomic data of the SARS-CoV-2 VOC 202012/01 and 501Y.V2 variants has been shared by the national authorities and uploaded to the Global Initiative on Sharing Avian Influenza Data […] and genomic surveillance of the virus continues, globally. [EV10]

5. Sources to corroborate the impact

EV1: WHO Statement on end of Ebola flare-up in Sierra Leone, World Health Organization, 17 March 2016 [Accessed 26 February 2021]

EV2: Email from Dhamari Naidoo, Laboratory Technical Officer, WHO Health Emergencies Program, World Health Organization [Dated 23 September 2018]

EV3: WHO situation reports detailing use of real-time sequencing to infer chains of transmission: Ebola Situation Report – 21 October 2015 & Ebola Situation Report – 30 March 2016

EV4: Clinical care for survivors of Ebola virus disease – Interim Guidance, World Health Organization, 11 April 2016

EV5: WHO R&D Blueprint, Public consultation - Pathogen genetic sequence data (GSD) [Accessed 26 February 2021] and Pathogen Data Sharing Code of Conduct

EV6: Example use of approach for other disease (i.e. rabies): Brunker K, Jaswant G, Thumbi SM et al. Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes [version 2; peer review: 3 approved]. Wellcome Open Res 2020, 5:3 DOI: 10.12688/wellcomeopenres.15518.2

EV7: Investigation of novel SARS-CoV-2 variant, Public Health England Technical briefing 1, 21 December 2020 & Technical briefing 2, 28 December 2020

EV8: (i) HC Deb 14 December 2020 vol 686 cc 23-47 – Covid-19 update, minutes of debate & (ii) Prime Minister's statement on coronavirus (COVID-19): 19 December 2020 [Accessed 26 February 2021]

EV9: Letter to the Vice Chancellor of University of Birmingham from COG-UK [Dated 24th March 2021]

EV10: WHO Disease Outbreak News: SARS-CoV-2 variants, 31 December 2020