

4: The World vs E.Coli

Researchers joining forces to sequence bacterial genomes

Affecting **16 countries** and almost **4,000 people**, the 2011 E.Coli outbreak took a terrible toll on Europe and beyond. The strain caused illness and deaths between May and June 2011, along with **massive financial losses** due to the danger of exporting contaminated produce.

The bacterial strain responsible for the outbreak was identified as O104:H4 by the Robert Koch Institute. However, scientists were looking for further information to understand where it originated. When the Beijing Genomics Institute openly **released the full genome sequence** of the outbreak strain, researchers from all over the world started to investigate it further. They managed to discover **ten isolates of the strain** from the outbreak thanks to a collaboration between universities, public, and private laboratories.

The investigation of the outbreak strain

Initially, researchers and public officials thought that the vehicle of the infection was the consumption of raw vegetables and salads. These were thought to be infected with a typical **Shiga toxin-producing E. coli (STEC)**, thus, public authorities recommended that people avoid eating certain types of food. Further collaborative **analyses** proved that the strain was, instead, very rare and managed to explain the symptoms of the diseases and complications reported by the nearly overwhelmed hospitals. The Robert Koch institute confirmed that German sprouts were

the source of the outbreak and a later assessment implicated Egyptian seeds that had been shipped to Germany in 2009. Interestingly, this conclusion was possible thanks to the combination of traditional and more recent methods of investigation, i.e., epidemiological methods and genome sequencing.

A literal example of “sharing is caring”

Understanding how the outbreak arose and where it originated was a very difficult task, but it was performed in record time. This is mostly because the datasets from sequencing initiatives were **publicly released in a timely manner**, so as to allow effective data re-use. The results from the analyses were covered via blogs, Twitter, and other websites, outside the traditionally slower scientific publication route. The combination between the analyses allowed the characterisation of the outbreak, which would have been extremely burdensome for a single organisation to undertake. On **GitHub**, 17 researchers gathered the results of their **crowdsourcing effort** to better understand the E.coli strain and compiled a list of 12 academic publications that arose from their work. Their crowdsourcing efforts were covered by 8 media outlets and showed that collaborations can increase a researcher’s profile and reputation while also serving a higher purpose such as analysing a bacterial outbreak. In addition, this collaboration showed that, when data is managed properly, newly-found highly virulent food-borne pathogens can be promptly analysed and tracked to their origin, thus, improving the safety of all citizens.

Title	4: The World vs E.Coli
Subtitle	Researchers joining forces to sequence bacterial genomes
Abstract	During the 2011 E.coli outbreak, the release of the bacterial genome in the public domain allowed researchers to reach conclusions quickly and effectively. The strain was tracked to a seed shipment from 2009 and the crowd-sourced analysis received high media attention.
Keywords	infection; e.coli; outbreak
Research subject area	MEDICAL AND HEALTH SCIENCES
Type of RDM impact/benefit	Efficiency in research and data re-use (e.g., reduce duplication of effort); Socio/Economic impact
Summary Impact Type	Impact on health and wellbeing
Facts and figures	17 authors working in a crowdsourcing effort 12 academic publications 8 media outlets covering the initiative
Original dataset from which the impact arose	
Maturity of the initiative/data source	Long-standing (5+ years)
Year (e.g., first data release, first output, year of impact)	2011
Organisations involved	BGI; Life Technologies; University Muenster; Health Protection Agency
Academic citations	
Links	http://blogs.nature.com/news/2011/06/the_german_e_coli_outbreak_40.html https://en.wikipedia.org/wiki/2011_Germany_E_coli_O104:H4_outbreak http://www.sciencemag.org/content/332/6035/1249 https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki https://www.cdc.gov/ecoli/general/ http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19890 http://www.dw.com/en/british-scientist-identifies-genetic-sequences-in-new-e-coli-strain/a-15133914