

# The use of long-read sequencing to investigate outbreaks of Shiga toxin-producing *E. coli* in humans.

Author: David R. Greig – Bioinformatician – UKHSA (UK Health Security Agency)

I think everyone can understand feeling unwell, a sore stomach and loose stools or diarrhoea. Now imagine, if you will, the same illness but so much worse, with terrible cramps, blood in your poo and even for some people, the need to go to hospital for surgery on their kidneys.

This awful illness is caused by a germ or as we scientists call it a bacterium, known as Shiga toxin-producing *E. coli*, this health professionals call STEC. STEC lives inside the stomach of farm animals like cows, sheep. People can catch bacteria from touching animals who carry STEC and by not washing their hands before touching the face or mouth or food. However, STEC is caught more commonly by eating contaminated food like salads, vegetables and even meat that has been in contact with another animal that has STEC.

Luckily STEC is rare compared to other illnesses, but for those who do catch it, the disease can be very nasty which is why we at UKHSA try to keep a very close look out for this particular bacterium. We do this is by receiving poo samples from sick patients infected with STEC. We find the STEC bacterium from within the samples, then we look at all the DNA from the bacterium. DNA is a kind of blueprint or code of the bacteria and what we do is read this code and compare it to all the DNA or code from all previous STEC sample that we have received. The more similar the code is between two bacteria the more likely the patients caught the bacterium from the same location or ate the same food that had the bacteria.

This current process of extracting and reading STEC DNA is takes a long time, about a week or so. My research aims to speed this up using the newest technologies which helps us read the DNA much faster and longer. Imagine a jigsaw broken up into a thousand tiny pieces and you don't know what the picture is, this would be like the existing method which reads tiny chunks of DNA. The new technology is like doing the same puzzle but with fewer and bigger pieces, so it is much easier to put back together even if we don't know what the picture is.

My research involves writing computer software to do the process of putting the jigsaw together on STEC DNA. This will reduce the time taken, instead of taking a week it can be done in two days. This means that we can find where the bacterium is coming from much faster and prevent more people from becoming ill with STEC. The other part of my research is to try to better understand what is contained in all the DNA of STEC, we call this the "genome". Completing the genome, i.e. putting the puzzle completely back together again, allows us to access information that would be otherwise be lost using the older methods. This information can be used to tell us how severe someone will get sick if they catch the bacteria or allows us to watch out for dangerous toxins that STEC can collect and make the illness more severe.