E. coli O104:H7 an example of evolution in foodborne pathogens and revolution in food microbiology

Prof Steve Forsythe

Introduction:

I have been holding back on releasing any comment until now due to the little accurate information that was initially available. This article was initially written on Saturday 4th June a couple of days after the draft genome of the infectious *E. coli* strain serotype O104 was released because I believe we have now passed a watershed in the investigation of foodborne pathogens. I will update these comments and observations periodically. Where possible I have used as reliable a source of information as possible and given URL links. I recognise that I have inferred the outbreak is due to food contamination, when in fact the vector as yet has not been identified. I have made this presumption based on previous outbreaks of the similar pathogenic *E. coli* serotype O157. However it is plausible that the organism was waterborne. The comments below cross-reference to sections in my book 'The Microbiology of Safe Food' (2nd Ed, Blackwell-Wiley).

The news in May to June 2011 is being dominated by the story of a severe, fatal outbreak of *E. coli* infections centred in Northern Germany. As the story continues to unfold it illustrates many aspects of modern and near future food microbiology. There is a rapid alert system across Europe called 'Rapid Alert System for Food and Feed' (RASFF: ec.europe.eu/food/food/rapidalert.index_en.htm). At the beginning of the outbreak in Germany cucumbers from Spain and Netherlands were found to contain enterohaemorrhagic *E. coli* and the authorities took action (See Fig 2 below). However it was later found that those pathogenic *E. coli* did not match the outbreak strain, nevertheless by this time the adverse publicity had caused additional product withdrawals. What appears to have been overlooked in the story is that acting on the consignment of cucumbers containing enterohaemorrhagic *E. coli* may have prevented a second severe outbreak.

Updates:

5th June: Contamination of beansprouts are named as the new suspected cause. There may have been poor hygiene either at a farm, in transit, or in a shop or food outlet. A restaurant in the northern port city of Lübeck is a possible place where the bacterium had been passed to humans. At least 17 people infected with *E coli* had eaten there. However as of 20:29 5th June the isolates from beansprouts have not been confirmed as matching those from the human cases.

6th June: Tests so far have not confirmed beansprouts as source. This is very confusing for the general public , and leads to a lack of confidence in the investigation. However outbreak sources can be linked two ways; epidemiological and by microbiology. The situation is that the beansprouts are linked 'epidemiologically' by analysis of questionnaires of affected people and looking for links such as common foods, or places. The link was made to people eating beansprouts at a restaurant. HOWEVER, we do not know if the beansprouts being tested are (a) the same batch as were supplied to the restaurant, and (b) from the restaurant or the farm. Since the outbreak has been going on for several weeks the chance of testing beansprouts from the same batch supplied to the restaurant is getting smaller ever day. So it is quite plausible that this highly publicized outbreak will end up as

epidemiologically linked to beansprouts, but not confirmed by microbiology. This often happens in outbreak investigations.

10th June: The Germany authorities have concluded that beansprouts from an organic farm in the northern village of Bienenbuettel were most likely the vector of the *E. coli* O104:H4 outbreak. However this is primarily based on epidemiology and the microbiological evidence is more circumstantial. Apparently the initial questioning of victims was very inadequate with beansprouts being dismissed too early in the investigation, and only re-considered recently. This consequently increased the human exposure period and number of infections as well as decreasing the chance of finding the relevant samples and isolates for microbiological analysis. The various questionnaires used can be accessed here http://www.rki.de/cln_178/nn_217400/EN/Home/ehec__Studien-Frageboegen.html.

July, 2011.

- Further investigations have traced the beansprouts to seeds from Egypt purchased in 2009. These also caused a smaller outbreak in France; EFSA Technical report; Fig 1.
- A number of papers have been published online describing the genomic analysis of *E. coli* O104 and its origins. The principle sequencing methods were Ion Torrent and Illumina; Fig 6 & 7.

Questions:

1: What is *E. coli* O104:H4? The code number '104' refers to the somatic antigen also known as the 'O' (not 0, zero) antigen, which vary according to the order and type of sugars making up the lipopolysaccharide (LPS) in the outer membrane. The '4' refers to the 'H' antigen type, which is due to the amino acids in the flagellum. Therefore we know the organism is motile, and can be distinguished from other varieties of *E. coli* using serotyping. For more details see sections 2.2.3 and 2.2.4.

2: How does the pathogenicity relate to the serotype? As explained above 'O104:H4' describes the surface antigens, it does not describe the virulence as such. Nevertheless the two have been inherited together and currently it is quicker to detect the surface antigens than virulence genes. However the situation is moving very fast and more direct detection of the virulence will undoubtedly be available very soon. See information on pathogenic *E. coli* in Section 4.3.3.

3: What is making this strain so virulent? Initial studies by the Robert Koch Institute (http://www.rki.de/EN/Home/homepage__node.html) using PCR methods showed the strain had the following gene profile: shigatoxin 1 negative, shigatoxin 2 (vtx2a) positive, intimin (eae) negative, enterohemolysin negative. It also contains the EaggEC virulence plasmid which was aatA positive (ABC-transporter protein gene), aggR positive (master regulator gene of Vir-plasmid genes), aap-positive (secreted protein dispersin gene), aggA positive (AAF/I-fimbrial subunit-gene) and aggC positive (AAF/I-fimbrial operon-gene). The organism has the ability to attach to the cells lining the

intestinal tract, and from there invade the human body by passing through the intestinal wall. It produces a variety of toxins which damage the kidney cells. See related information in Section 4.3.3.

4: What about antibiotic resistance? The following antibiogram was obtained from the Robert Koch Institute (http://www.rki.de/EN/Home/homepage__node.html) :

Resistant: Ampicillin,amoxicillin/clavulanic acid, piperacillin/sulbactam, piperacillin/tazobactam (AES VITEK), cefuroxim, cefuroxim-axetil, cefoxitin, cefotaxim, cetfazidim, cefpodoxim, streptomycin, nalidixinsäure, tetracycline, and trimethoprim/sulfamethoxazol. Extended spectrum beta-lactamases (ESBL): CTX-M-15 positive, other ß-lactamases: TEM-1 positive.

Sensitive: Imipenem, meropenem, amikacin, gentamicin, kanamycin, tobramycin, ciprofloxacin, norfloxacin, nitrofurantoin, chloramphenicol, fosfomycin, and nitrofurantoin.

5: Does the organism infect females more than males? Yes. Looking at the data from the second *Eurosurveillance* article (2nd June, 2011) it shows the male:female ratio for all age groups (Fig 3). It is notable that the ratio is skewed to greater female infections even for pre-school children. The only age group where the incidence is almost equal is in the age range 10-14 yr. The predominance of infections in females is highly unusual and at the current time is unexplained.

6: Why were cucumbers named initially? There is a rapid alert system across Europe called 'Rapid Alert System for Food and Feed' (RASFF; http://ec.europe.eu/food/food/rapidalert.index_en.htm). At the time of the outbreak in Germany cucumbers from Spain and Netherlands were found to contain enterohaemorrhagic *E. coli* and the authorities took action (See Fig 2). However it was later found that those pathogenic *E. coli* did not match the outbreak strain. Nevertheless by this time the adverse publicity had caused additional product withdrawals. What appears to have been overlooked is that acting on the consignment of cucumbers containing enterohaemorrhagic *E. coli* may have prevented a second severe outbreak. See sections 1.10 and 1.12.6 for information on RASFF and the cost of foodborne infections.

7: How can *E. coli* O104:H4 be detected? This depends on your starting material. It is easier to isolate a bacterium from a normally sterile site such as blood than food. In food microbiology the target organism may be stressed due to the food processing, and also in a mixed culture. For *E. coli* O104:H4 the situation is more difficult as other types of Enterobacteriaceae and *E. coli* may be present in a food sample. The problem is that one does not pick off and identify in detail every single colony from selective agars such as MacConkey and VRBGA. The same problem arose when *E. coli* O157:H7 came to prominence, but in time new agars and detection methods were developed which were based on phenotypic (ie. sugar fermentation) and physiological differences which could be used to differentiate the O157:H7 serotype strains from other *E. coli* strains. See section 5.2.1.

8: What is 'STEC' and 'VTEC' and are they not all *E. coli* O157:H7 anyway? STEC stands for 'shiga-toxigenic *E. coli*' and VTEC stands for 'verotoxin producing *E. coli*'. These terms are essentially interchangeable. VTEC was the initial term used before it was realised that the damage to vero cells (tissue culture cell line) was due to the same toxin as found in *Shigella*. These are groups of pathogenic *E. coli* strains which have acquired additional virulence genes from the closely related bacterium *Shigella*. *E. coli* O157:H7 is only one variety of STEC/VTEC. Others include O111 and O26. See Section 4.3.3 for more detail.

9: How is an outbreak investigated? First one has to recognise that hindsight can be very cruel. If you know the organism causing the infection then one has an idea of the common sources to look into. Obviously if the source turns out to be untypical then one is criticised for being too blinkered. Outbreak sources can be linked two ways; epidemiological investigation and by microbiology. In the epidemiology investigation, questionnaires are completed by the victims regarding recent eating habits and travel abroad. This may involve matched controls. If a likely common factor is eating at a restaurant then analysing foods for the target bacterium can be carried out. The problem in Germany appears to be the delay in the authorities being informed. Doctors may not report promptly to the local authorities, who have a week to inform the state authorities who then have a week to inform the relevant investigative group (Robert Koch Institute). So three weeks may have elapsed before data has been gathered, and any microbiological analysis started. It will be necessary to not only match food or water isolates at the species level, but actually DNA fingerprinting (Question 10 below). There are therefore many opportunities for infections to be missed, and not investigated. The chances of finding the same batch of food in a long food distribution chain gets less each day. Also accurate recall of food eaten several days before is not totally reliable and prone to assumption. If you ate a salad what are you likely to recall? Most likely 'lettuce, tomatoes and cucumber', but what about the rest of the garnish? Hence the questionnaire results will be biased.

10: What about using pulsed-field gel electrophoresis (PFGE) for tracing the source of the outbreak? Yes, PFGE is commonly used for tracing outbreaks of organisms including *E. coli* O157 and *Salmonella*. The PFGE profile has been obtained (Fig 5) and will be of use in the current investigation. For further information on PFGE and PulseNET see sections 1.12.3 and 5.5.1.

11: What was the point of sequencing the genome? From the answer to Question 8 we have already seen that *E. coli* can acquire new DNA sequences encoding for virulence genes from other bacteria. It appears that on this occasion an enteroaggregative strain of *E. coli* (EAggEC) has acquired the shiga toxin (normally seen in another *E. coli* variety; STEC or EHEC) via a lambda bacteriophage (process called 'transfection'). A similar strain was isolated from a 6 year old girl in 2001 and partially sequenced. The transferance of DNA is commonly known as 'horizontal gene transfer' when it occurs between two different species, but this time it was intraspecies, that is between two types of *E. coli*.

The technology for DNA sequencing has improved so quickly that in contrast to the initial *E. coli* O157:H7 outbreaks in USA (1993), Japan (1996), and Scotland (1996) it is almost possible to sequence the whole bacterium in the same amount of time as running PFGE. However even using the 2nd and 3rd generation DNA sequencers such as Ion Torrent (Life Technologies) and Illumina® one needs skilled bioinformaticians to put the short DNA reads together into the correct contigs. Please note that to date the *E. coli* O104:H4 in Genbank is an unfinished genome as the contigs have not been closed, ie regions in-between have not been sequenced. However one can carry out DNA sequence alignments and comparisons to other pathogenic varieties of *E. coli*. Fig 5 is a comparison with enteroaggregative *E. coli* (EAggEC) and is courtesy of Yongmei Li (Life Technologies). The Genbank accession number for the Ion Torrent sequenced genome is AFOB01000000. It is this rapid sequencing of the strain which has revealed the combination of *E. coli* traits which have occurred to generate the 2011 outbreak *E. coli* O104:H4 strain. How this strain has acquired additional genes compared with previous clinical isolates of *E. coli* O104:H4 strains remains to be discovered, but for now this new variant is the focus. For further reading on the evolution of *E. coli* and genomic analysis

see sections 2.3.2, 2.9.1 and 2.9.5 (Fig 2.19), as well as articles by Mellmann *et al*. (2011), and Rhodde *et al*. (2011). See also Fig. 8 at the end of this article.

Further analysis combining Ion Torrent and Illumina sequencing were published online in July (2011) and can be downloaded; Rhodde *et al.* (2011) and Mellmann *et al.* (2011). There are interesting aspects on the origin of *E. coli* O104 genome which have been revealed very quickly.

Summary.

There are two activities in foodborne pathogen outbreak investigations; epidemiology and microbiology. The former is based on responses to questionnaires by the victims and investigation of the hygienic practices on the food production process. The later (microbiology) requires laboratory analysis with methods which are specific and sensitive to the target organism. In this case a subvariety of a common bacterium, *E. coli*. It requires isolates from the same batch of suspect food as ingested by the victims. When there is a long period between initial cases and food sampling then it is less likely to have the same batch for analysis and the evidence becomes more circumstantial.

The advances in DNA sequencing have enabled the rapid analysis of the emergence of the highly virulent strain of *E. coli* O104:H4 erupting in Northern Germany. This will hopefully lead to improved detection and surveillance in the future and also inform us of the selection forces driving its evolution. The previous high profile outbreaks of *E. coli* O157 such as in Scotland led to changes and reinforcement in HACCP approaches to food safety and hygiene training. While witnessing the evolution of *E. coli* we are also experiencing a revolution in the investigation of foodborne infections. Nevertheless, speed in acquiring isolates from victims and food vector(s) are still a major issue.

Useful sources of information:

EFSA Task force investigation report: http://www.efsa.europa.eu/en/supporting/pub/176e.htm
European Centre for Disease Prevention and Control (ECDC): http://www.ecdc.europa.eu
Eurosurveillance: http://www.eurosurveillance.org/
Federal Bureau of Risk Assessment (BfR): http://www.bfr.bund.de/en/home.html

Fig.1. Trace back of E. coli O104:H4 to seeds from Egypt. Screen grab from EFSA Task force report.

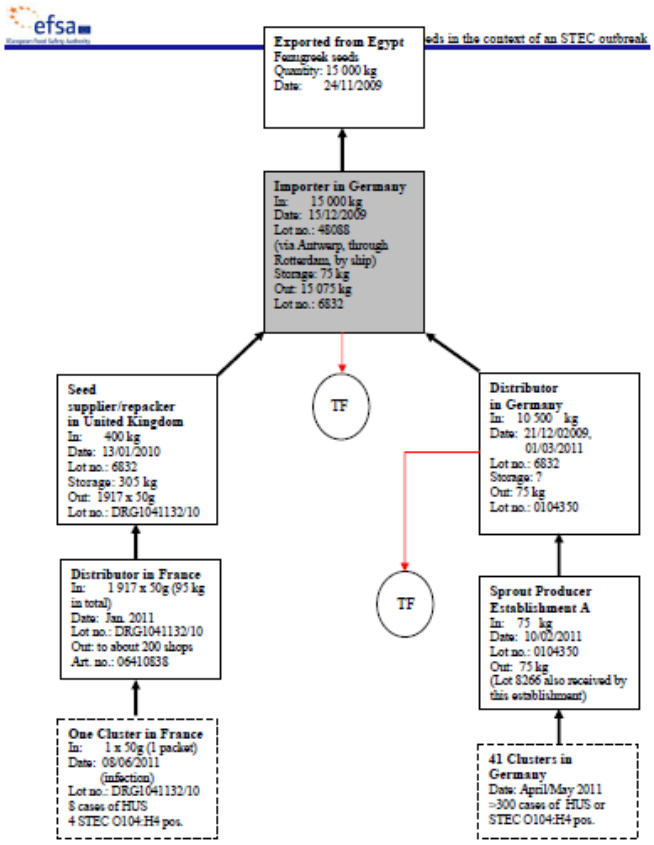
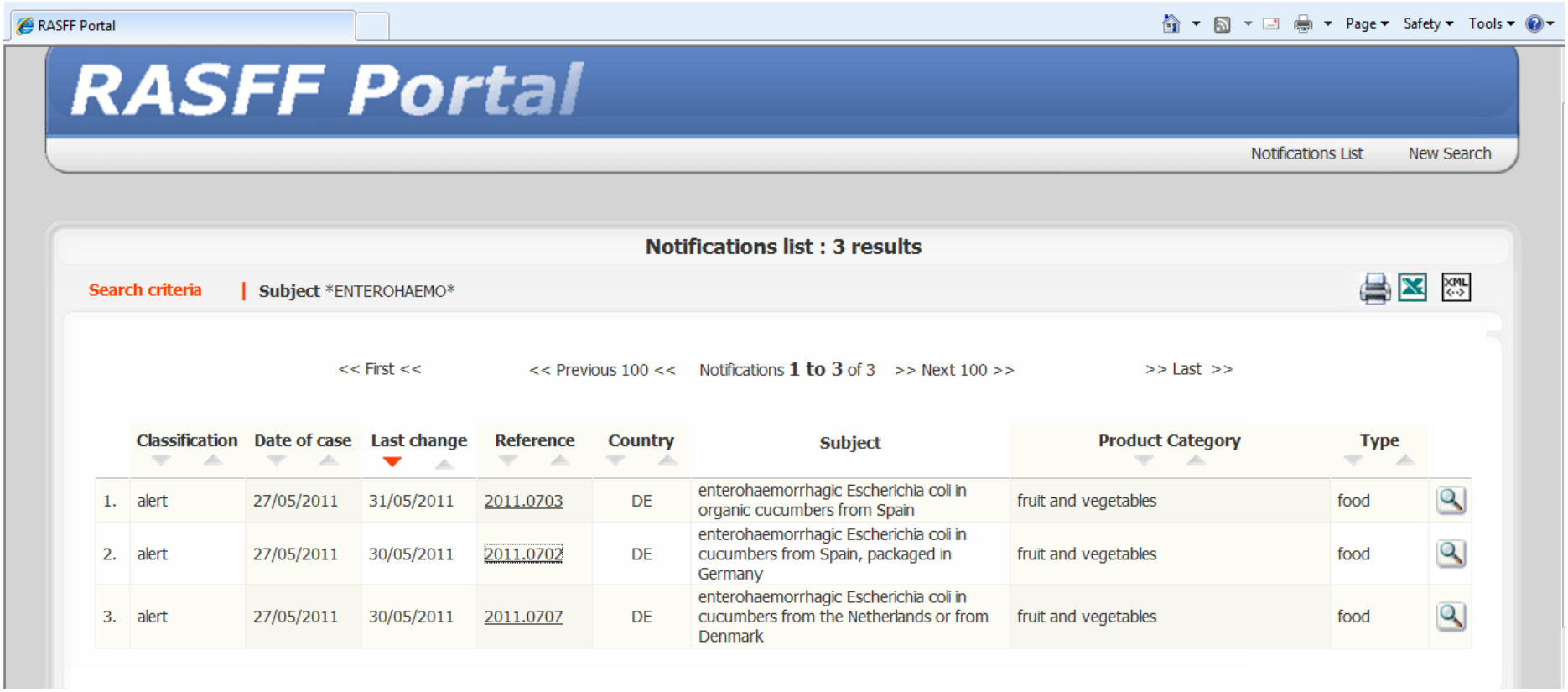


Figure 1: Network graphic showing the trace back (black thick arrows) (incl. lot no., quantities and dates) of fenugreek seeds lot no. 48088 from the two clusters in Germany and France via all identified distributors/producers to the company in Egypt. The red thin arrows indicate the ongoing tracing forward work done and are discussed in chapter 3.2.
TF=Trace forward, see section 3.2

Fig 2. RASFF portal alert news showing the detection of enterohaemorragic *E. coli* in cucumbers at the end of May 2011.



	Classification	Date of case	Last change	Reference	Country	Subject	Product Category	Type
1.	alert	27/05/2011	31/05/2011	2011.0703	DE	enterohaemorrhagic Escherichia coli in organic cucumbers from Spain	fruit and vegetables	food
2.	alert	27/05/2011	30/05/2011	2011.0702	DE	enterohaemorrhagic Escherichia coli in cucumbers from Spain, packaged in Germany	fruit and vegetables	food
3.	alert	27/05/2011	30/05/2011	2011.0707	DE	enterohaemorrhagic Escherichia coli in cucumbers from the Netherlands or from Denmark	fruit and vegetables	food

Fig 3. Cumulative incidence of HUS cases in Germany according to age and sex. Source: *Eurosurveillance* Vol. 16, Issue 22, 02 June 2011.

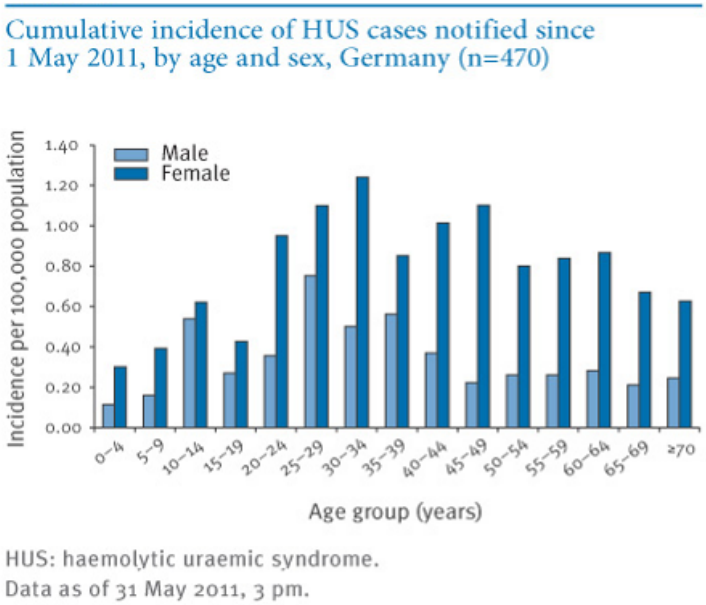


Fig 4. *Xba*I restriction digest of *E. coli* O104:H4 lanes 2-4 (three separate isolates), *Salmonella* marker in lanes 1 & 5. Source Robert Koch Institute, http://www.rki.de/EN/Home/homepage__node.html.

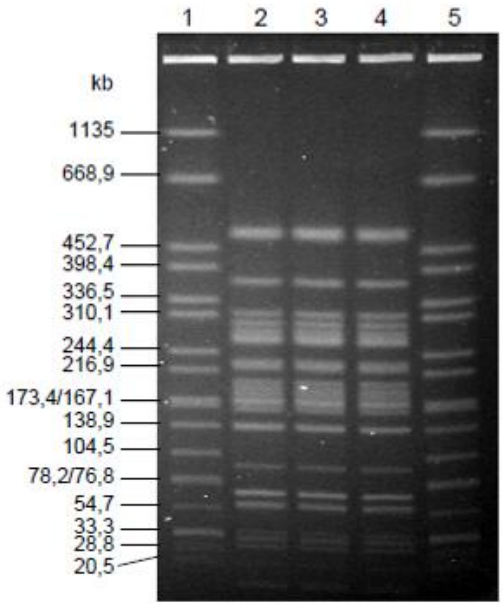


Fig 5. Genome:genome comparison of *E. coli* O104:H4 (upper strand) with enteroaggregative *E. coli* (lower strand) using the program Mauve. Image courtesy of Yongmei Li, Life Technologies.

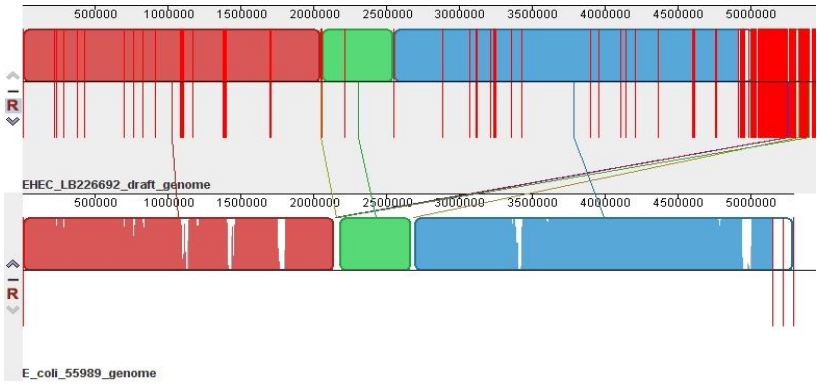


Fig 6. Timeline of *E. coli* O105 genome sequencing. Screen grab from Rhodde *et al.* 2011 New Eng J Medicine.

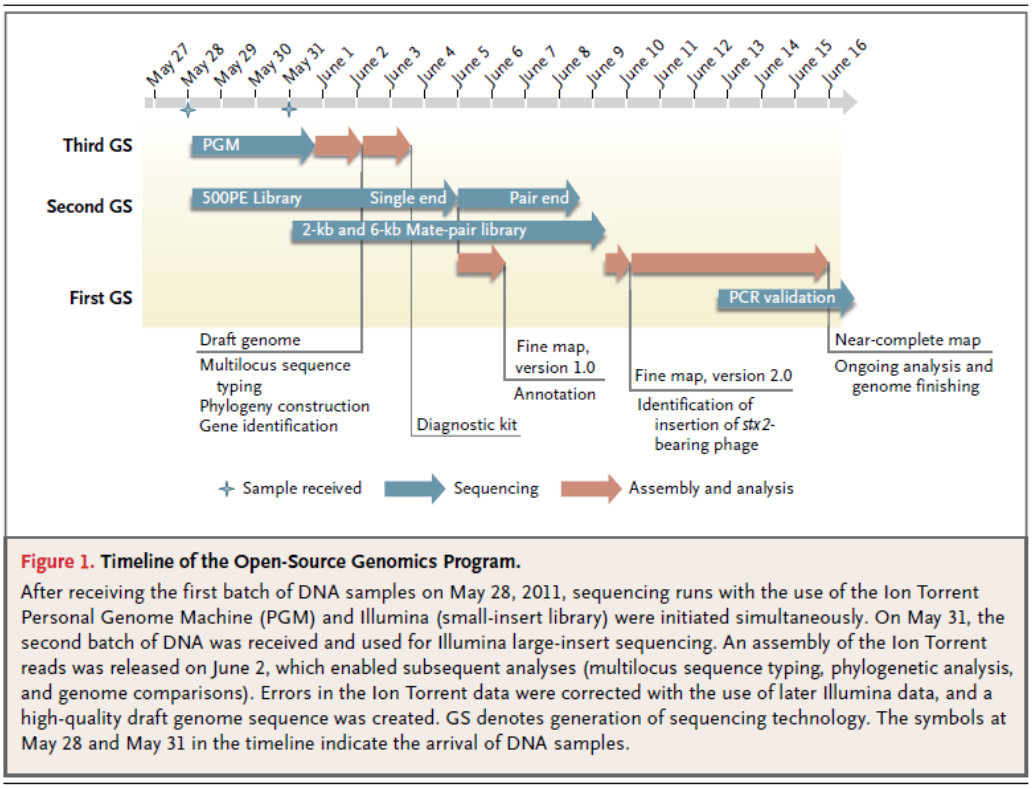


Figure 1. Timeline of the Open-Source Genomics Program.

After receiving the first batch of DNA samples on May 28, 2011, sequencing runs with the use of the Ion Torrent Personal Genome Machine (PGM) and Illumina (small-insert library) were initiated simultaneously. On May 31, the second batch of DNA was received and used for Illumina large-insert sequencing. An assembly of the Ion Torrent reads was released on June 2, which enabled subsequent analyses (multilocus sequence typing, phylogenetic analysis, and genome comparisons). Errors in the Ion Torrent data were corrected with the use of later Illumina data, and a high-quality draft genome sequence was created. GS denotes generation of sequencing technology. The symbols at May 28 and May 31 in the timeline indicate the arrival of DNA samples.

Fig. 7 Time line of *E. coli* O104 outbreak and analysis. Screen grab from Mellmann et al. (2011), PLoS ONE.

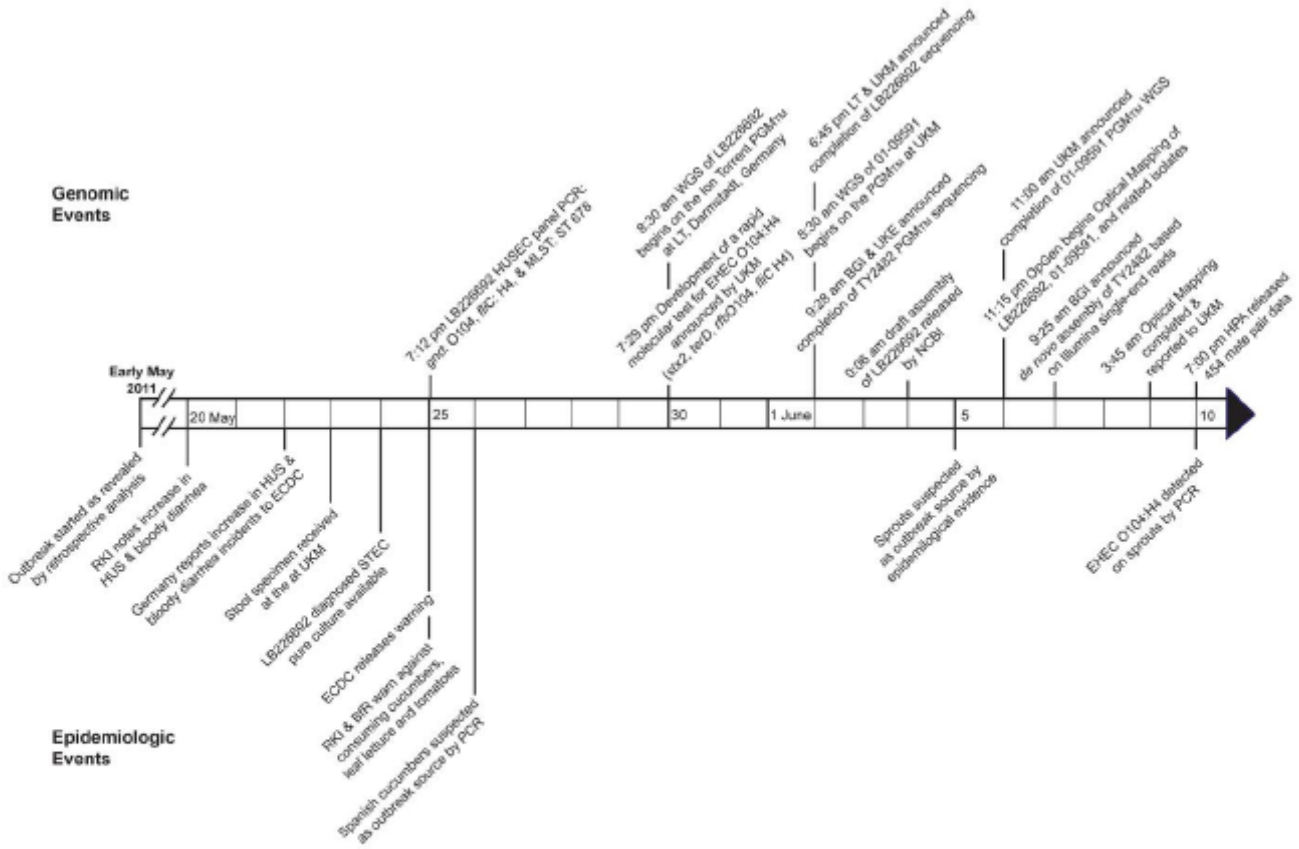


Figure 1. Events timeline of German EHEC O104:H4 outbreak. Major events relating to the outbreak epidemiology (below arrow) and those relating to genomic elucidation efforts (above arrow) are noted separately in the graph. Lines within the arrow indicate single day progression, with the date noted every 5th day. Events span from early May 2011 to early June 2011. Times are noted in Central European Time (CET). Abbreviations: BfR = Bundesinstitut für Risikobewertung (Federal Institute for Risk Assessment, Germany), BGI = Beijing Genomics Institute (People's Republic of China), ECDC = European Center for Disease Prevention and Control (Sweden), HPA = Health Protection Agency (United Kingdom), HUS = hemolytic uremic syndrome, LT = Life Technologies Group, PGM™ = Ion Torrent Personal Genome Machine™, RKI = Robert Koch Institute (Germany), ST = multilocus sequence type, UKE = University Hospital Hamburg (Germany), UKM = University Hospital Muenster (Germany), WGS = whole genome sequencing.
doi:10.1371/journal.pone.0022751.g001

Fig. 8. Phylogenetic analysis of *E. coli* O014 and relationship to other varieties of pathogenic *E. coli*. Screen grab from Mellmann *et al.* (2011).

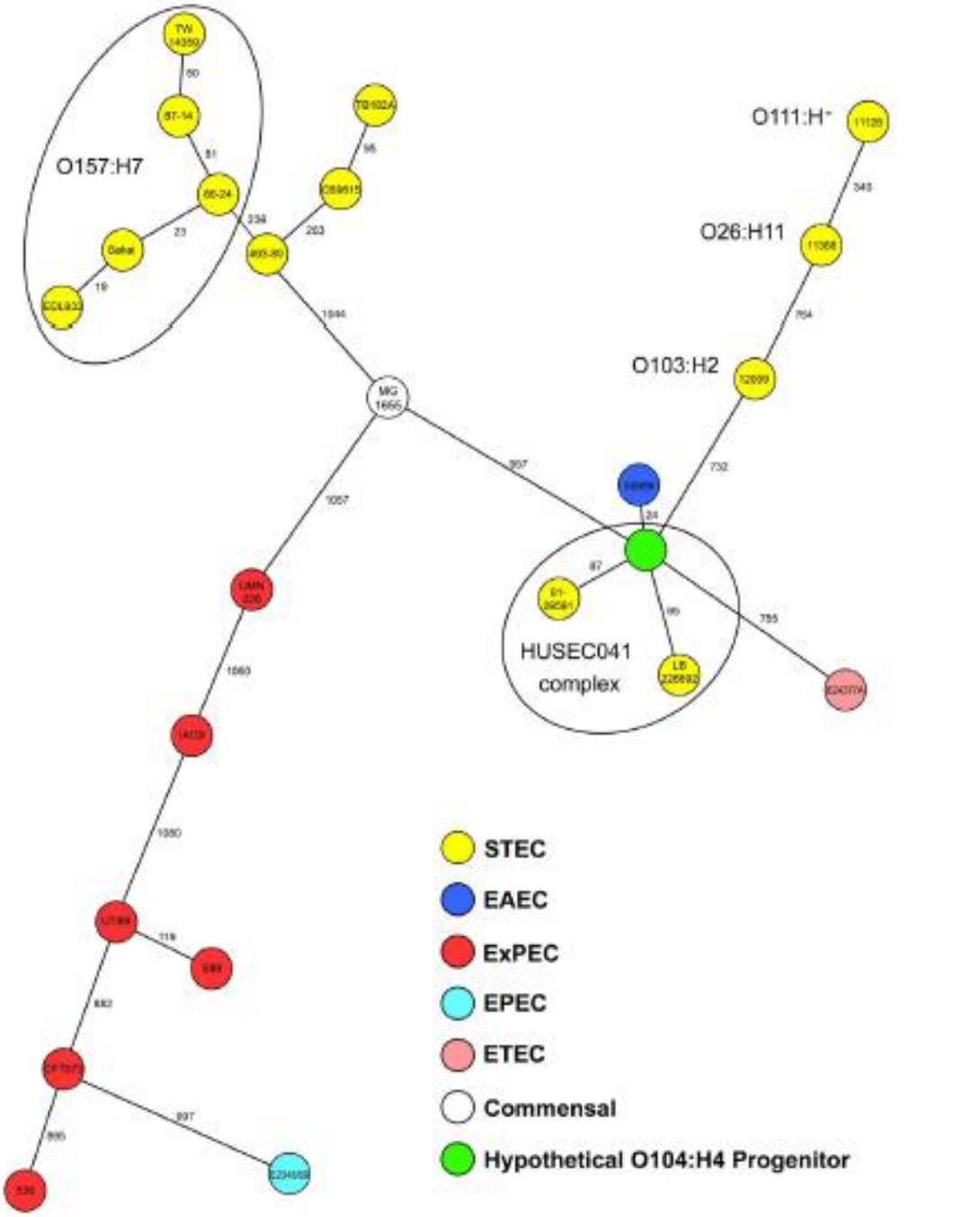


Figure 3. Phylogentic placement of German EHEC O104:H4 outbreak strain. Minimum-spanning tree based on allelic profiles of *E. coli* core genome genes (n = 1,144) portraying the phylogenetic relationship of the EHEC O104:H4 outbreak strain (LB226692), the historical EHEC 01-09591 (HUSEC041), additional *E. coli* strains representing the most common EHEC serotypes, intestinal and extraintestinal *E. coli* pathovars and commensals, from the NCBI RefSeq database. In addition, an *in silico* generated hypothetical O104:H4 progenitor is included. Each dot represents an allelic profile, the number on connecting lines represent the number of alleles that differ between two profiles. The different pathovars (EHEC, EAEC, ExPEC, EPEC, ETEC, commensals) are defined by colors and the EHEC serotypes are indicated.
doi:10.1371/journal.pone.0022751.g003