Large scale expansion of mobile elements in specific hotspot regions of the German outbreak *Escherichia coli* O104:H4

Lisa C Crossman

The Genome Analysis Centre, Norwich Research Park, Norwich, NR4 7UH

Abstract

The outbreak strain of *E. coli* O104:H4 has been sequenced by several groups and made available publically for CrowdSourcing purposes. Genome comparisons of the complete finished sequence of TY2482 (BGI), have been made with the draft assemblies of c22711 (PacBio) and an historical outbreak strain from 2001 (U. Münster & Life Technologies). A plasmid, pTY2, carrying the *agg* operon specifying the genes for aggregative adherence fimbriae reveals a frameshift in a gene, *aggB*, that may result in altered binding *in vivo* compared to the unframeshifted state. Comparisons additionally reveal the presence of genomic islands specific to the outbreak strain relative to other sequenced *E. coli* strains. These regions, and islands shared with the closest previously sequenced relative *E. coli* 55989 have been analysed for insertion sequences and transposable elements. Several islands found in the above strains that are not present in other sequenced *E. coli* are found to harbour a large-scale expansion of mobile elements that are by and large confined to these hotspot or permissive areas of the chromosome. The implication is that these regions are in genomic flux and may represent specific areas of future concern due to the possibility of mobilisation of the associated genomic features likely responsible for the pathogenic features and antibiotic resistance seen in these strains.

Introduction

In May 2011, Germany reported an outbreak of a severe food poisoning caused by an *E. coli* strain[1]. Specific features of this strain included the ability to cause Haemolytic ureamia syndrome (HUS) symptoms with bloody diarrhoea. The strain was serotyped as the O104:H4 type, rarely seen in food poisoning. Numbers of serious food poisoning cases quickly reached epidemic proportions. Those affected by this outbreak were mainly adults, approximately 60% were female, rather than the more common population demographic of infants, elderly and the immunocompromised. The number of affected people presenting with HUS in a short time scale was described as unusual[1]. Salad vegetables were initially identified as the carrier foodstuff, although the exact food source was later pinpointed as beansprouts produced on factory scale. However, it proved difficult to find traces of the bacterium at the source farm. The strain was later discovered on beansprouts in the refuse of a family that had become ill. Previous reports have implicated raw sprouts in food poisoning epidemics[2,3].

BGI China initially sequenced the outbreak strain *E. coli* TY2482, by the Ion Torrent and Illumina platforms[4]. The sequence was provided in draft form to the scientific community as a CrowdSourcing project[5]. BGI subsequently finished the assembly to completion.

CrowdSourcing researchers[6,7] found that the strain is most closely related to the *E. coli* 55989, a strain originally isolated in Central Africa some years ago and implicated in severe diarrhoea there. However, the outbreak *E. coli* has gained a phage conferring the ability to make the Shiga toxin[8,1].

The outbreak strains have been designated EAHEC, with the suggestion that they are close relatives of or derived from EAEC that have gained the ability to cause haemolytic uraemia by means of obtaining the Shiga-toxin producing bacteriophage, but do not carry an EHEC locus of enterocyte attachment and effacement (LEE)[9,10]

In total, 11 sequences of outbreak isolates have been made publically available as shown below[6].

TY2482 (BGI in collaboration with University Medical Centre Hamburg-Eppendorf)[4]
LB226692 (Life Tech in-house in collaboration with University of Münster)[11]
H112180280 (released earlier with a 454 scaffold) plus 4 additional isolates (Health Protection Agency, Colindale, UK)
Two isolates, GOS1 and GOS2 (Göttingen Genomics Lab, Germany)[10]
C22711 sequenced by Pacific Biosciences in addition to several related strains[11]
Historical HUSEC041 O104:H4 isolate (MLST type ST678) from 2001 has also been sequenced by Univ. Hospital Münster and Life Tech[11].

An outbreak of food poisoning associated with HUS was later reported from France[12]. This outbreak was initially attributed to sprouted seeds served at a school fete. The source of both German and French outbreaks are attributed to the consumption of raw sprouted seed products[3]. The seeds have become contaminated by causes unknown.

In this investigation, genome comparisons of the above publically available sequences TY2482, c22711 and an outbreak strain of the HUSEC041 clone from 2001 have been carried out to identify sequences unique to these strains that may explain the unusual pathogenic features. Particular genomic variable islands have been identified which carry a massive expansion of IS and transposable elements highly confined to these regions in the sequences of the outbreak strain and of its closest sequenced relative, the EAEC, *E. coli* 55989.

Methods and Materials

Annotation was examined using the Artemis software[13]. IS and transposases were characterised using Fasta and BlastP against the non-redundant sequence databases. Each prospective mobile element sequence was also examined with Blastn against the IS element database, ISFinder[14].

Results

1.1 Chromosome comparisons

E. coli strains were compared using reciprocal best match analysis. The predicted proteins were plotted as a circular diagram against orthologous genes from related strains as shown in Figure 1.

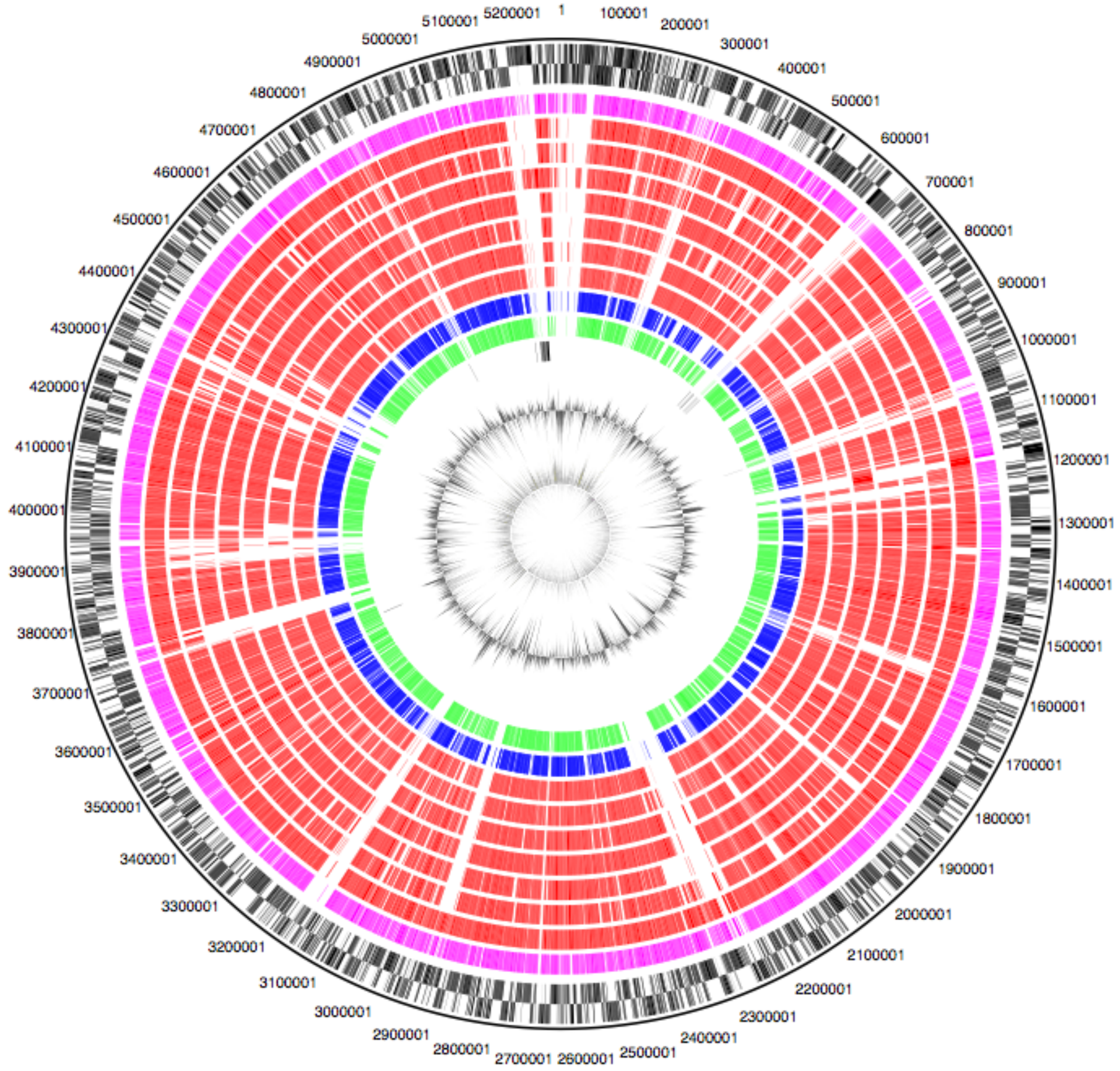

Figure 1

Circular figure comparing the *E. coli* TY2482 chromosome against related organisms. Circles from outermost to innermost represent the following: DNA coordinates from TY2482 chromosome, coding sequences (CDS) from TY2482. The magenta circle shows orthologous genes from the historical 2001 isolate. Red circles show the orthologous genes from the following *E. coli* strains: 55989, CFT073, O157:H7, 24377A, O83H1, K12DH10B, HS, respectively. The green circle represents *Salmonella* Heidelberg whilst the blue circle represents *Shigella flexneri*. Black CDS matches at the top and two o'clock of the figure are matches to STEC phage. The innermost circle shows a GC content plot.

The plasmids were similarly compared using reciprocal best match analysis. The two major plasmids, pTY1 and pTY2 were concatenated (laid end-to-end) for this analysis to allow for the possibility of plasmid rearrangements that would create matches across the two major plasmids to one single plasmid from another organism.

The results of this analysis are shown in Figure 2.

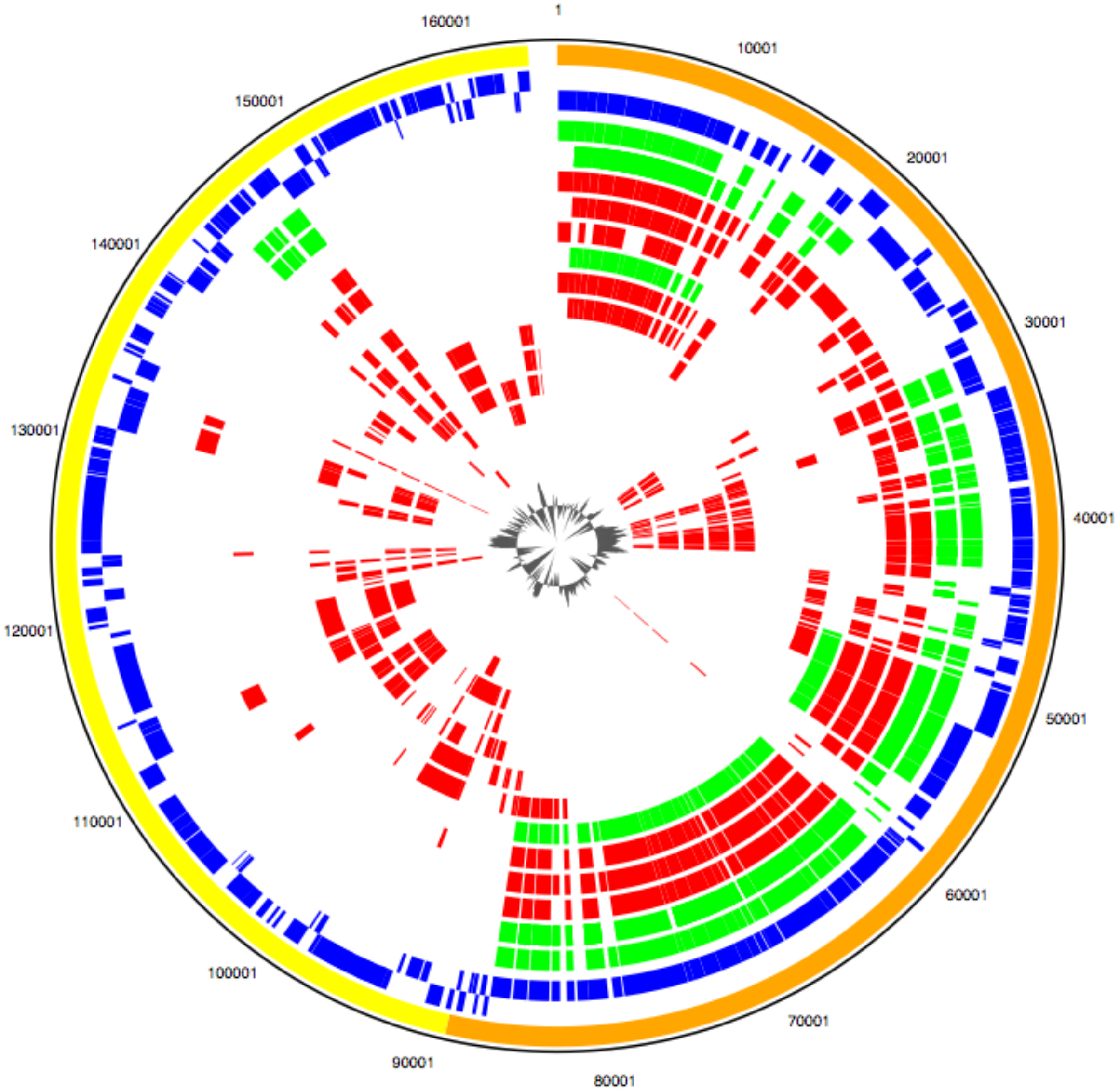

Figure 2

The sequences of pTY1 and pTY2 were laid end-to-end and used as a reference against orthologous genes from other plasmid genomes in order to check for cross plasmid matches. From outermost to innermost the circles are as follows: 1. DNA coordinates for the concatenated pseudomolecule. 2. Orange represents pTY1 and yellow represents pTY2. 3. Royal blue indicates the predicted coding sequences of pTY1 and pTY2. 4. Green circles represent matches with a *Salmonella* plasmid, red matches are with *E. coli* plasmids. Orthologous genes are shown on a single strand for clarity. Green circle 1 = *Salmonella* Heidelburg, 2= *Salmonella* Kentucky, 3 = pEC_Bactec, 4 = pEK204, 5 = pECOED, 6 = *Salmonella* R64, 7 = pCoo, 8 = p746, 9 = p55989, 10 = pO86A1, 11 = pAA, 12 = p666, 13 = pEK499, 14 = pEK516, 15 = GC content (%).

In total, 15 major variable regions (VRs) were identified in the chromosomes of the outbreak strains TY2482, c22711, outbreak 2001 and 55989 relative to other sequenced *E. coli* strains with particular reference to the *E. coli* commensal strain HS[15,16]. These regions have been systematically numbered with their major constituent coding sequences shown in Table 1.

Table 1. Summary of major Variable Regions (VRs)

Region	Approximate coordinates (TY2482 sequence)	Summary of constituents	Possible derivation
VR1	1-65686	Phage	phage
VR2	280616-314076	Phage	phage
VR3	620670-681089	Phage	phage
VR4	812447-850662	Phage	phage
VR5	1009409-1057971	Phage	phage
VR6	1152832-unclear	Phage	phage
VR7	1602015-1630327	Phage	phage
VR8	1902693-1924756	Hypotheticals and GTP binding factor from phage	unclear
VR9	2262828-2351785	Type VI secretion proteins	mobile
VR10	2868642-2904821	Arsenic resistance, peptidases	mobile
VR11	3114201-3162558	Mobile elements and antibiotic resistance	mobile
VR12	3698419-3748155	Aerobactin siderophore	mobile
VR13	3883383-3949691	Sugar transporters and synthesis	mobile
VR14	4326836-4355752	Region with strong similarity to that from VR9	mobile
VR15	5167132-5228810	Phage	phage

Mobile = mobile element

Variable region (VR) 15 carries the phage *stx* genes, to which are attributed the production of Shiga toxin. This toxin is of known pathogenicity and has been described as the cause of haemolytic uraemia (HUS), which can lead to kidney failure[17].

From these variable regions, on examination six showed a large expansion of mobile elements and partial mobile elements. The VRs may occur at hotspot or permissive regions. Regions which have putative defined integration sites include VR3, integrated into anaerobic DMSO reductase chain B; VR8, *smpB*; VR9, tRNA Phe ,VR11, tRNA Sel; VR13, tRNA Leu; VR14, tRNA Thr.

Mobile element distribution was examined in *E. coli* strains TY2482, 55989 and the commensal *E. coli* strain HS. Whilst each isolate may show variability in mobile elements present, mobile elements were found strongly clustered to specific islands in TY2482 and 55989. In the commensal *E. coli* HS[15,16], 87 IS elements were identified by the previously described methods, and these were far less clustered than in TY2482 and 55989.

VR11 is truly unique to the 2011 outbreak strain. In particular, VR11 possesses 15 partial or complete insertion sequence (IS) elements as compared to a total of 103 identified from the annotation, BLAST and ISFinder across the entire chromosome. For a region the length of VR11 at 48 358 bp, the expected amount of insertion sequences should the number 103 be randomly distributed across the chromosome is 0.92. Hence it is clear that there is a large expansion of mobile elements in this region. VR11 additionally carries a large number of antibiotic resistance gene determinants that are separated from each other either by partial or complete mobile elements. The region may have originally derived from an extra chromosomal element or integrated plasmid since both a replication protein and a partial *repA* gene are present. The GC content of VR11 is a high 55.8% rather than the overall chromosomal composition of 50.8%. Supposing the GC content tends to the average for a genome over time according to stabilising selection, taken together with the lack of this region found in other related strains, this is highly suggestive of a recent transfer event. VR11 has interrupted a tRNA, a known indicator of a hotspot region for DNA integration[18,19]. The tRNA shares closest matches to that for *selC*, selenocysteine. Integrations in this region have been previously described in other *E.coli* strains[19].

VR11 partially shares features as a chromosomal element with *Salmonella enterica subsp. enterica serovar* Heidelberg. Figure 3 shows the BLASTn matches of VR11 as compared to the equivalent region from *Salmonella* Heidelberg.

Figure 3.

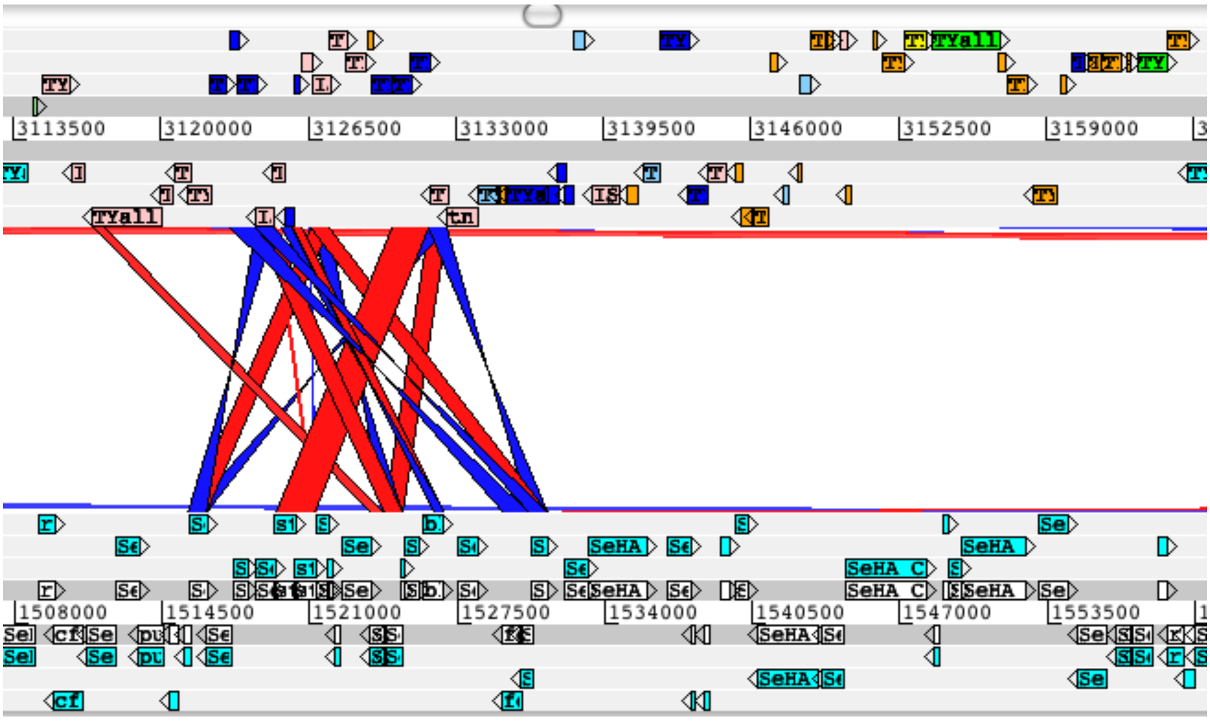

Figure 3.

Red bars indicate regions of forward matches of VR11 to the *Salmonella* Heidelberg sequence. Blue bars indicate reverse matches. The TY2482 sequence is shown at the top of the figure whilst the *Salmonella* Heidelberg sequence is shown at the base of the figure. Annotated genes are shown as coloured boxes on the TY2482 sequence. Pink boxes indicate mobile elements and royal blue boxes indicate antibiotic resistance genes.

The N terminal region of VR11 shares the closest sequence matches with *Salmonella* sequences and *Salmonella* plasmids, whereas the C-terminal region shares the closest sequence identity with other *E. coli* sequences and *Shigella flexneri*.

The antibiotic and other resistances putatively encoded by VR11 include sulphonamide, beta lactams, two streptomycin or other aminoglycosides, mercury, tetracycline and a drug permease. Also present is an autotransporter and a haemolysin expression modulating protein as well as YeeV-YeeU toxin-antitoxin system that targets other bacterial cells.

Further resistances in the genome such as heavy metal and third generation cephalosporins are present elsewhere.

1.2 Plasmid comparisons

1.2.1 Plasmid pTY1

The plasmid here denoted pTY1 (CTX-*bla*-TEM, IncI1 type plasmid), 88 695 bp, is conjugative and additionally possesses a shufflon of the type found in *Salmonella* plasmid R64 and plasmid p746 from *E. coli* strain p1392[20,16]. This shufflon acts to alter the tip of the conjugative pilus protein, *pilV* to allow attachment and the passing of genetic material to different strains of *E. coli* and relatives[20]. This plasmid also encodes beta lactamase resistance. There are 4 putative mobile element genes present on this plasmid.

1.2.2 Plasmid pTY2

Plasmid pTY2 (IncFIIA/FIIB type), 75 330 bp, carries the *agg* operon involved in the production of the aggregative adherence fimbriae and also carries 47 partial or complete mobile elements.

The distribution of the mobile elements on this plasmid as compared to those on the chromosome in terms of the particular IS element families present are shown in Figure 4. The distribution of families is not the same as those from the chromosome, suggesting a horizontal transfer event such as conjugation has occurred from a related organism.

Figure 4.

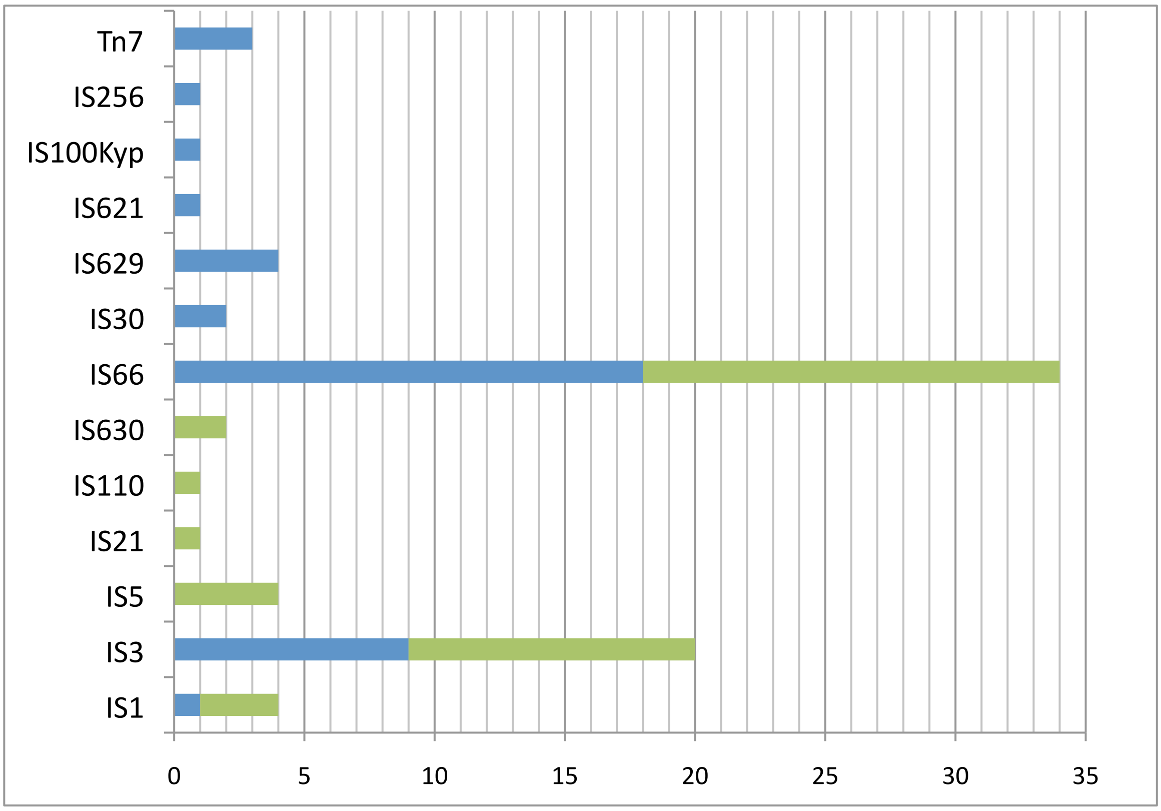

Figure 4. Families of IS elements located to the chromosome and the plasmid pTY2. Green indicates IS element families located to pTY2, whilst blue indicates IS element families located to the chromosome.
The profile of the IS element families differs except with respect to IS66 and IS3 families. This is suggestive of a lateral gene transfer event for the pTY2 plasmid, which is in keeping with our knowledge of the history of this strain.

Manual annotation of the *agg* operon of pTY2 from TY2482 indicates a frameshift centrally in the *aggB* gene (Figure 5). The gene product of *aggB* was previously described as non-essential for aggregation, however, it may alter the properties of aggregative binding[21]. This frameshift is also present in many of the reads of the c22711 sequence from PacBio, which may be indicative of its *in situ* presence in the outbreak *E. coli* strains.

Figure 5A

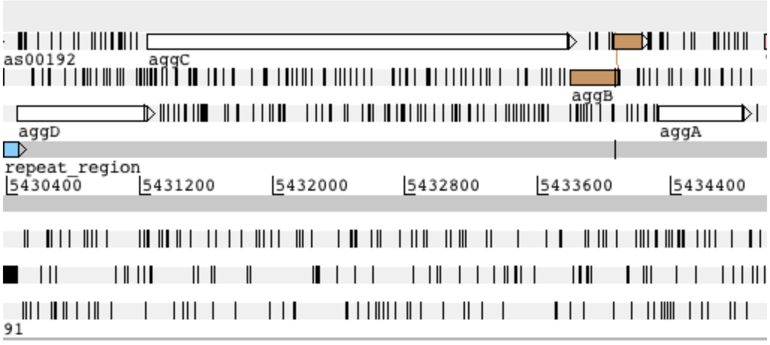

Figure 5B

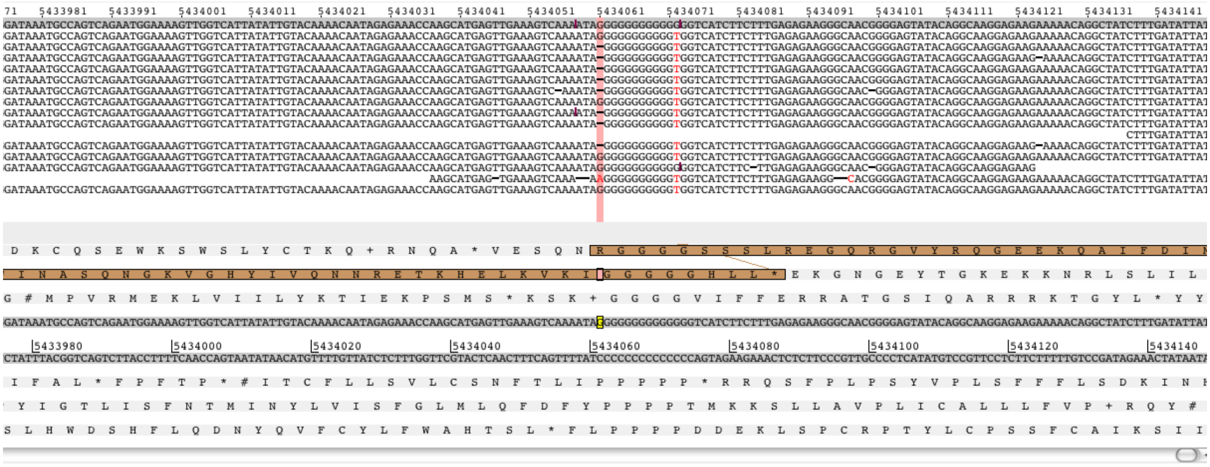

Figure 5A

An Artemis view of the annotation of the *agg* operon from pTY2. The *agg* operon of pTY2 showing the *aggB* gene with a putative frameshift mutation.

Figure 5B

Reads from the c22711 PacBio seqence overlaid on pTY2 sequence. The c22711 strain possesses a SNP G->T relative to the TY2482 pTY2 sequence. A long homopolymeric run of guanines are situated adjacent to the potential frameshift in *aggB*. Approximately half of the reads from c22711 have an extra G leading to a frameshift like the sequence of TY2482. Examination of the reads in this region from two of the Illumina sequenced isolates from HPA, Colindale, UK shows that most although not all of the reads support a frameshift in *aggB*, with an additional synonymous SNP T->G located just downstream from the frameshift present in some of the reads. It could be possible that both the frameshifted and non-frameshifted forms of this gene are present in subsets of cells of these outbreak strains along with the SNP.

Also carried on this plasmid are the error repair genes *impAB* and both a complete and an interrupted copy of the IgA1 protease gene. This protease has previously been implicated as a virulence determinant[22].

1.2.3 Plasmid pTY3

Plasmid pTY3 appears to be a 1 549 bp entirely selfish DNA plasmid[6,23], highly related to plasmid pO26-S1 previously isolated from the Shiga toxin producing *E. coli* O26:H11 strain H30. The sole identifiable gene on this sequence is *repA* and no mobile elements are found on this plasmid. Such short selfish DNA plasmids have been seen previously as *E. coli* pKL1[24] and the *Bacteriovorax marinus* plasmid pBMS1[25].

Discussion

The genome sequence of the outbreak *E. coli* O104:H4 was examined very rapidly by CrowdSourcing analyses. In future this may represent a method to look at genomes of medically relevant bacteria in a very timely fashion. The main features of a genome may be elucidated rapidly by CrowdSourcing means, however, there is still plenty of scope for more in-depth analysis that could reveal substantial additional clues to prevent further outbreaks.

The genome sequence of the outbreak *E. coli* O104:H4 putatively represents a means to look at an evolutionary event that has very recently occurred. From the severe and recent symptoms not previously before seen, we must infer that the genome has in some way mutated or picked up DNA from another source by transfection, transduction or conjugation. An alternative explanation that clearly does not appear to be the case would be that of zoonosis.

Potentially transferred regions containing a large expansion of mobile elements may have previously resided in a related strain that was subject to high amounts of transposition or mobile element activity, possibly as a result of exposure to environmental stressors. As a group EHEC are known to carry large amounts of mobile elements and phage DNA[26]. The outbreak strain has been described as EAHEC, hence the mobile element distribution is called into question. From the results of this study, the number of mobile elements is not vast, and they are particularly found associated with specific integrated regions such as VR11.

The outbreak strain also harbours three plasmids, pTY3 is 'selfish' containing only replication apparatus, pTY1 is an IncI1 type *bla*-CTX-M15 that bears strong resemblance to other IncI1 *E. coli* plasmids carrying *bla* genes for beta-lactam antibiotic resistance and a conjugative apparatus with a shufflon system similar to that seen in *Salmonella* and enterotoxigenic *E. coli* H10407. It is possible that this apparatus can also enhance the ability of the organism to attach to surfaces. The pTY2 plasmid is known as an enteroaggregative (Eagg) plasmid, however, it shows significant differences to other plasmids of this type, including the plasmid from *E. coli* 55989.
The *agg* operon shows a frameshift in the *aggB* gene. Mutations in this gene have previously been suggested to result in altered binding to targets. Sequencing reads from different NGS platforms show the presence of both frameshift and non-frameshift reads in addition to the presence or absence of a nearby synonymous SNP G->T. It is possible that both types could co-exist in cell populations.

A major finding from this study is that the distribution of families of IS elements found on pTY2 are not the same as the families of IS elements on the chromosome, which points to pTY2 having previously been harboured by an alternate strain and gained by TY2482 by a horizontal gene transfer event such as conjugation.

A timeline for genomic events is suggested in Figure 6.

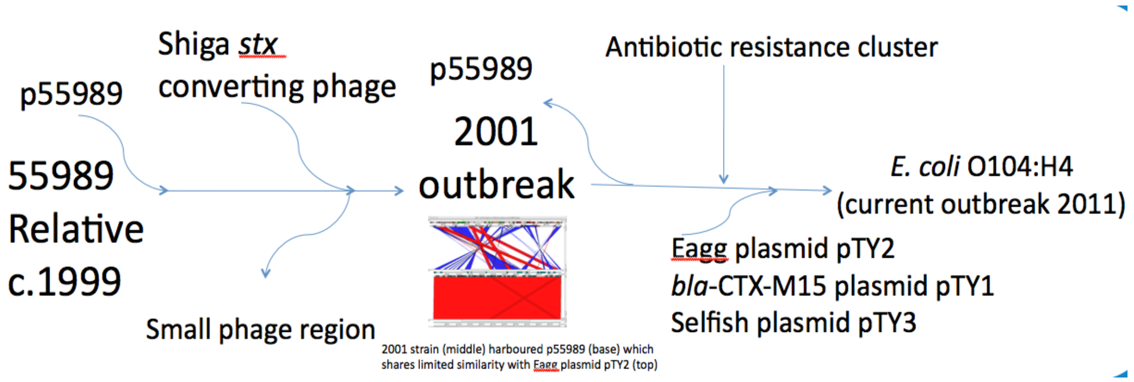

Figure 6

E. coli 55989 was isolated *c.* 1999. A converting phage was gained in a relative of this EAEC, and a small phage region was lost to form the 2001 outbreak strain. Post 2001, a p55989-like plasmid was lost and three other plasmids were gained in addition to antibiotic resistance gene cluster, VR11.

Chromosomal regions that are truly unique to this strain include phage-derived regions and those associated with mobile elements. In addition, the vast expansion of mobile elements on the plasmid pTY2 and in vital specific and regions unique to this strain may indicate an evolutionary strategy for rapid change for this organism. These inserted regions likely reside in specific hotspots for foreign DNA insertion.

It is not known whether the expansion of mobile elements in this strain could explain its high virulence or what allowed it to cause such a devastating foodborne outbreak, however, it could be a factor in increased virulence, gaining a foothold and in particular in gaining antibiotic resistance genes. Whilst treatment with antibiotics may be contraindicated in outbreak cases involving the Shiga toxin, in some cases antibiotic treatment has been used. In addition, recent transfer events involving antibiotic resistance are always a cause for concern.

Prediction of future food poisoning epidemics are not currently possible, however, a strain carrying many virulence determinants such as the genetic backbone of the current outbreak strain would seem to be at a high risk of gaining sequences that will allow it to cause a severe outbreak.

Acknowledgements

This study was supported by the BBSRC, South Norfolk Council, Norwich City Council, Norfolk County Council, the East of England Development Agency (EEDA) and the Greater Norwich Development Partnership. LC thanks TGAC for support and thanks the following CrowdSourcing researchers in order of joining: Marina Manrique, Raquel Tobes, Nick Loman, David Studholme, Konrad Paszkiewicz, Torsten Seeman, Simon Gladman, Kat Holt, Pablo Pareja Tobes, Tim Dallman, Anthony Underwood, Raju Misra, Patrik D'haeseleer, Kwan Lab, Shrinivasro Mane, Scott Edmonds and Nico Petty.

LC thanks the BGI (China), University Medical Centre Hamburg-Eppendorf, The Health Protection Agency (HPA, Colindale), Göttingen Genomics Lab, Germany, Pacific Biosciences, University of Münster and Life Technologies for making sequences publically available.

References

1. Frank, C. *et al.* Epidemic Profile of Shiga-Toxin-Producing *Escherichia coli* O104:H4 Outbreak in Germany - Preliminary Report. *The New England Journal of Medicine* (2011).doi:10.1056/NEJMoa1106483
2. Taormina, P.J., Beuchat, L.R. & Slutsker, L. Infections associated with eating seed sprouts: an international concern. *Emerging Infect. Dis.* **5**, 626-634 (1999).
3. Pennington, H. *Escherichia coli* O104, Germany 2011. *The Lancet Infectious Diseases* **11**, 652-653 (2011).
4. Rohde, H. *et al.* Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N. Engl. J. Med.* **365**, 718-724 (2011).

5. Kupferschmidt, K. Scientists Rush to Study Genome of Lethal *E. coli. Science* **332**, 1249 -1250 (2011).
6. Manrique, M. & Tobes, R.at <https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki>
7. Cheung, M.K., Cheung, M.K., Li, L., Nong, W. & Kwan, H.S. 2011 German *Escherichia coli* O104:H4 outbreak: Alignment-free whole-genome phylogeny by feature frequency profiles. *Nature Precedings* (2011).doi:10.1038/npre.2011.6109.2
8. Cheung, M.K., Cheung, M.K., Li, L., Nong, W. & Kwan, H.S. 2011 German *Escherichia coli* outbreak: Prophage analysis of close-assembled TY2482 against 55989 using PHAST. *Nature Precedings* (2011).doi:10.1038/npre.2011.6110.1
9. Denamur, E. The 2011 Shiga toxin-producing *Escherichia coli* O104:H4 German outbreak: a lesson in genomic plasticity. *Clin. Microbiol. Infect.* **17**, 1124-1125 (2011).
10. Brzuszkiewicz, E. *et al.* Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Archives of Microbiology* (2011).doi:10.1007/s00203-011-0725-6
11. Mellmann, A. *et al.* Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE* **6**, e22751 (2011).
12. Gault, G. *et al.* Outbreak of haemolytic uraemic syndrome and bloody diarrhoea due to *Escherichia coli* O104:H4, south-west France, June 2011. *Euro Surveill.* **16**, (2011).
13. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).
14. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32-36 (2006).
15. Rasko, D.A. *et al.* The pangenome structure of *Escherichia coli*: comparative genomic analysis of E. coli commensal and pathogenic isolates. *J. Bacteriol.* **190**, 6881-6893 (2008).
16. Crossman, L.C. *et al.* A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *J. Bacteriol.* **192**, 5822-5831 (2010).
17. Bielaszewska, M. *et al.* Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *The Lancet Infectious Diseases* **11**, 671-676 (2011).
18. Chen, N. *et al.* The pheV phenylalanine tRNA gene *Klebsiella pneumoniae* clinical isolates is an integration hotspot for possible niche-adaptation genomic islands. *Curr. Microbiol.* **60**, 210-216 (2010).
19. Vejborg, R.M., Hancock, V., Petersen, A.M., Krogfelt, K.A. & Klemm, P. Comparative genomics of *Escherichia coli* isolated from patients with inflammatory bowel disease. *BMC Genomics* **12**, 316 (2011).
20. Komano, T., Kim, S.R. & Yoshida, T. Mating variation by DNA inversions of shufflon in plasmid R64. *Adv. Biophys.* **31**, 181-193 (1995).
21. Savarino, S.J., Fox, P., Deng, Y. & Nataro, J.P. Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. *J. Bacteriol.* **176**, 4949-4957 (1994).
22. Halter, R., Pohlner, J. & Meyer, T.F. IgA protease of *Neisseria gonorrhoeae*: isolation and characterization of the gene and its extracellular product. *EMBO J.* **3**, 1595-1601 (1984).
23. BGI assembly of German *E. coli* outbreak strain « bacpathgenomics. at <http://bacpathgenomics.wordpress.com/2011/06/18/bgi-assembly-of-german-e-coli-outbreak-strain/>
24. Burian, J., Guller, L., Macor, M. & Kay, W.W. Small cryptic plasmids of multiplasmid, clinical *Escherichia coli*. *Plasmid* **37**, 2-14 (1997).

25. Crossman, L.C. *et al.* The parting of the delta proteobacterial ways; A small core of common genes illuminates similarities in the predatory lives of the highly divergent marine *Bacteriovorax marinus* SJ and the terrestrial *Bdellovibrio bacteriovorus*. **Submitted**, (2011).
26. Ogura, Y. *et al.* Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17939-17944 (2009).