life technologies™

# Analysis to assay

## Rapid characterization of shiga toxin–producing *Escherichia coli*

A recent outbreak of shiga toxin–producing *Escherichia coli* had serious health implications, causing the deaths of at least 22 individuals in northern Germany as of early June 2011 [1]. This particular isolate can cause hemolytic-uremic syndrome (HUS), a type of kidney failure, which inevitably increases the mortality rate in affected populations.

Multiple teams across the globe have been analyzing this outbreak with the ultimate goals of rapidly generating a complete genomic sequence, annotating the sequence, elucidating the reasons for high mortality, and, most pressingly, developing a specific assay for detecting the strain in food samples.

With a broad portfolio of assays and a plurality of sequencing solutions at its disposal, Life Technologies was uniquely positioned to assist with the development of a complete workflow solution—from initial sequence elucidation to an isolate-specific assay. In collaboration with multiple clinicians and public health centers, Life Technologies scientists deployed complementary technologies and associated bioinformatics to speed the development of an assay. The robust strain characterization workflow is outlined in Figure 1.

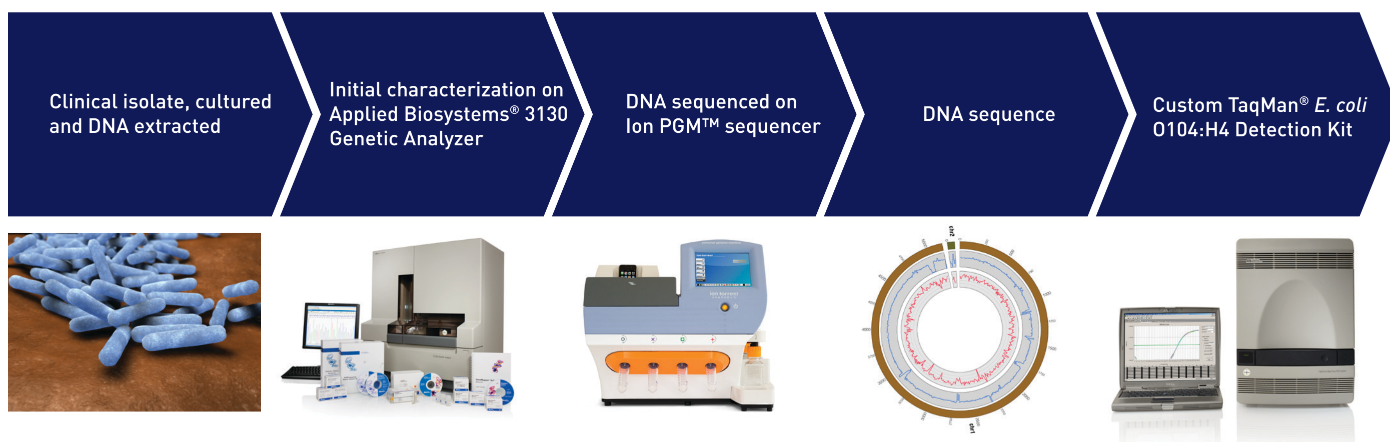
**Robust initial characterization**

The National Consulting Laboratory on Hemolytic Uremic Syndrome in Münster, Germany, initially characterized the infectious agent using classical serotyping and Sanger-based capillary electrophoresis (CE) sequencing. They performed Multi-Locus Sequence Typing (MLST) [3] using an Applied Biosystems® 3130 Genetic Analyzer. The MLST technique characterizes isolates using internal fragment sequences derived from several bacterial housekeeping genes. The fragments are approximately 450 to 500 base pairs in length, and the fragment strategy employed ensures that each gene is represented. For the target isolate, the alleles identified at these loci define the allelic profile or sequence type (ST). By comparing to a database of previously identified strains, a putative identification can be made [4]. The outbreak O104:H4 strain was identified as belonging to ST678, the same type as *E. coli* strain HUSEC41, an Stx2-positive O104:H4 strain isolated in 2001.

Subsequently, 800 bases of the Stx2 gene were also sequenced using Sanger (CE) sequencing, which led to the identification of the Stx2a allele. Using this initial information, the strain was characterized as *E. coli* O104:H4, *stx2*-positive, *eae*-negative, *iha*-positive, ESBL-positive, but gentamicin- and fluoroquinolone-susceptible [5].

In order to assist in dealing with the outbreak, Life Technologies was able to make available a previously designed TaqMan® based Stx2 assay, enabling identification of the Stx2 gene.

**Figure 1. Robust workflow for strain characterization and assay development, highlighting Life Technologies assays and instruments used in the process.**




**Box 1. Assembly of sequence reads from the Ion PGM™ sequencer.**

1. Align reads to the public *E. coli* 55989 genome [12] with TMAP.
2. Generate consensus sequence with SAMtools [13].
3. Split consensus at zero-coverage regions (perl script) to generate consensus contigs.
4. Using MIRA [14], map reads to consensus contigs; the remaining reads were used in a *de novo* assembly (to fill gaps and unique regions not present in the reference region).
5. Merge contigs with CAP3 [15].

Results are shown in Table 1.

**Table 1. Summary of results obtained from assembly of Ion PGM™ sequencer reads.**

| | |
|---|---|
| percent A | 24.7 |
| percent C | 25.4 |
| percent G | 25.3 |
| percent T | 24.7 |
| Sum contig length | 5,450,264 |
| Number of contigs | 364 |
| Mean contig length | 14,973 |
| Median contig length | 762 |
| N50 value | 181,540 |
| N90 value | 14,537 |
| Max contig length | 475,662 |

## Prompt genomic characterization

Even though the initial characterization of the strain was complete, developing a strain-specific assay required the organism's full genomic sequence. Such an assay would provide rapid answers regarding contaminated food and determine if *E. coli* O104 is actually present. A strain-specific assay would also assist in identifying the source of infection, important for both public health strategies and to minimize potential diplomatic complications [6, 7].

With the advent of next-generation sequencing, large amounts of sequence data can be generated very rapidly. For Ion Personal Genome Machine™ (PGM™) sequencer analysis, DNA is fragmented, and each discrete DNA construct is amplified on a single bead—a process termed "clonal amplification." This streamlined sample preparation of the PGM™ sequencing workflow is critical in such serious public health outbreaks. In fact, the Ion Personal Genome Machine™ sequencer allows researchers to generate high-quality sequence from a genome of this size in about 2 hours. Moreover, each individual run on an Ion 314™ chip generates at least 10 Mb of high-quality sequence. Combined with simple upstream sample preparation, this technology can deliver a complete research workflow—from DNA sample to sequence data—in a single day.

## Rapid sequence generation—the Ion PGM™ sequencer workflow

The purified bacterial DNA was fragmented by sonication (Covaris®, Inc.) into ~150 bp fragments. Alternative methods for fragmenting DNA such as mechanical or enzymatic shearing can also be employed to generate fragments of this size in a random process (*de novo* assembly assumes random fragmentation). The library was then constructed using the standard Ion Fragment Library Kit protocol [8]. Two libraries were constructed, each derived from 1 µg of DNA. One library had an average fragment size of 187 bases and the other library had an average fragment size of 211 bases (including adapters). Briefly, Ion-specific adapters were ligated to the DNA fragments, and subsequently amplified on beads, with each bead having the product of a single fragment amplified thousands of times. The beads were then loaded onto an Ion 314™ chip, and sequenced using semiconductor sequencing [9], generating ~100 base reads (Figure 2). Using two Ion PGM™ systems, ten Ion 314™ chips were run in two days. In order to assess sequence quality, internal controls were used.

The reads were then assembled using a combination of reference based– and *de novo*–assembly strategies [see Box 1, Figures 3 and 4]. The reads were also mapped to EDL933 (including plasmid) in order to assess throughput. Following sequence generation, the sequences were assembled (see Box 1), generating the results detailed in Table 1 with 18x average coverage. This *E. coli* O104:H4

strain LB226692 genome sequence was submitted to NCBI (accession number AFOB00000000.1) [10]. A second sequencing project was carried out using the Ion PGM™ sequencer by the Beijing Genome Institute (BGI) and is available at NCBI with accession number SRX067313 [11].

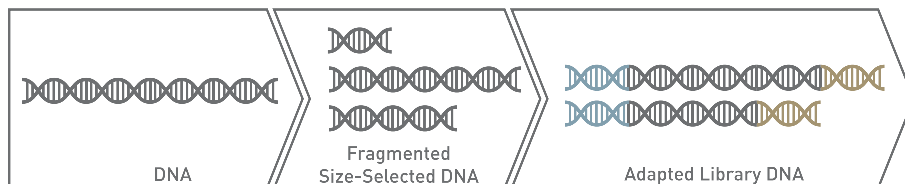
The true power of this process is demonstrated by the fact that this initial sequence generation was achieved in days both in Europe (Life Technologies, Darmstadt Training Center in collaboration with Münster University) and in China (BGI-Shenzen) [16]. Access to this sequence enabled the teams to rapidly pinpoint key features of the virulent strain. Dr. med. Alexander Mellmann, a scientist at the German National Consulting Laboratory for Hemolytic Uremic Syndrome (HUS) at the Institute of Hygiene, University Hospital Münster, remarked, "The rapid whole-genome sequencing results enabled us to discover within days a unique combination of virulence traits...and makes this German outbreak clone a unique hybrid of different *E. coli* pathovars" [17].

## Isolate-specific assays

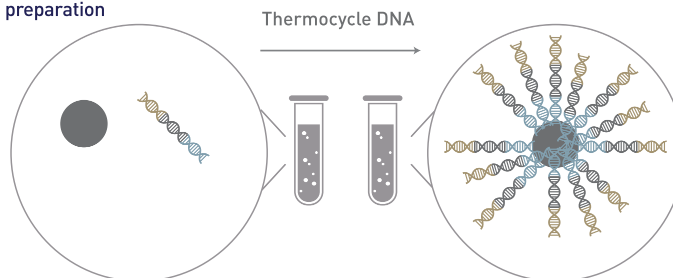
The National Consulting Laboratory on Hemolytic Uremic Syndrome (HUS) in Münster, Germany, developed a gel-based endpoint PCR assay using four primer pairs targeted to four different genes: stx2, terD, O104rfb, and fliC H4 [18]. By analyzing the size of the PCR products, the presence or absence of these genes can be assessed (the outbreak strain is predicted to be positive for all four genes). In order to make rapid analysis possible and increase testing throughput, Life Technologies scientists worked

**A. Ion PGM™ sequencer workflow**

**B. Library preparation**




**C. Template preparation**




**D. Sequencing and base calling**

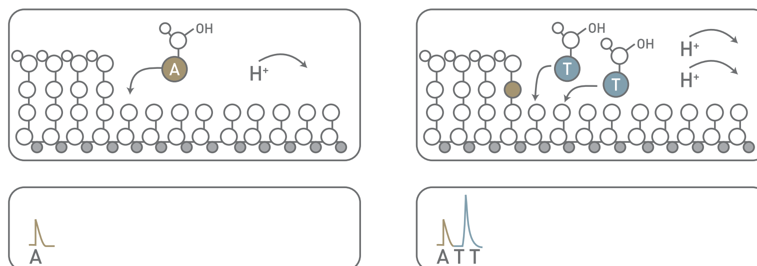



**Figure 2. Speed—the Ion Torrent Sequencing Workflow.** Ion sequencing workflow consists of preparation of a genomic DNA fragment library. Adaptors are ligated onto the ends of the fragments and clonally amplified onto Ion Sphere particles and enriched. These particles are then loaded into the Ion Chip and placed on the Ion PGM™ sequencer for sequencing.

```
CTGATGAAAAATATCGCCCGAGTTGAGAAACAACGCGAACTTGCCTTGTGCCATCGCAATGCCTTTGTTCATGGCGTCGTAGATACCGTTATCTGGCTCGCTGACAAAGCGTAGGTTAAAGATACCAT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CTGATGAAAAATATCGCCCGAGTTGAGAAACAACGCGAACTTGCCTTGTGCCATCGCAATGCCTTTGTTCATGGCGTCGTAGATACCGTTATCTGGCTCGCTGACAAAGCGTAGGTTAAAGATACCAT
```

**Figure 3: An example read from an Ion PGM™ sequencer used in the genome assembly (Figure 4) showing 128 base pair perfect read.**

**Box 2. TaqMan® Assay design process.**

1. Identify sequences to compare (ensure that true positive identification of target strain is achieved while true negative results are maintained for closely related strains).
2. Compare sequences bioinformatically; identify unique features of target strain using internal pipeline that discovers signature sequences using MUMmer [19].
3. Design TaqMan® Assay using Primer Express® software.
4. Validate against closely related strains and contaminants (e.g., if testing for a food pathogen, ensure that there is no response from the background matrix).

Note how the availability of sequence is an essential prerequisite for assay design.

to develop a real-time PCR assay to detect the *E. coli* O104:H4 strain in food samples (Box 2). The TaqMan® *E. coli* O104:H4 Detection Assay is ideal for food contamination detection, as it has two inherent levels of specificity—from binding of the PCR primers, and binding and subsequent hydrolysis of the TaqMan® probe (Figure 5). This allows for highly accurate identification of a virulent strain in foods without the confounding false positives that can arise from closely related strains. The TaqMan® assay approach is also exquisitely sensitive with a wide linear range (Figure 6).

Armed with the newly acquired reference sequence, scientists at Life Technologies executed a fast assay design process, which allowed an assay to be designed and tested within a week.

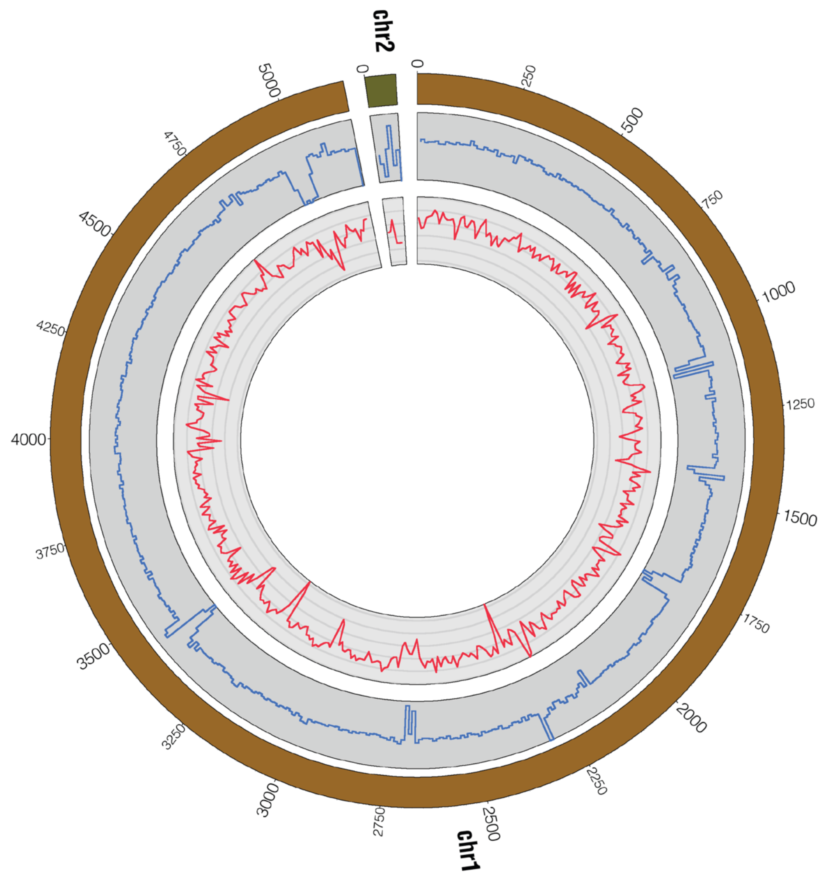



**Figure 4. Circos plot of genome assembly.** Sharp drops in coverage (blue line) represent differences between assembled sequence and the reference *E.coli* 55989.
chr1 = chromosome, chr2 = plasmid, red lines = GC content, blue lines = coverage.

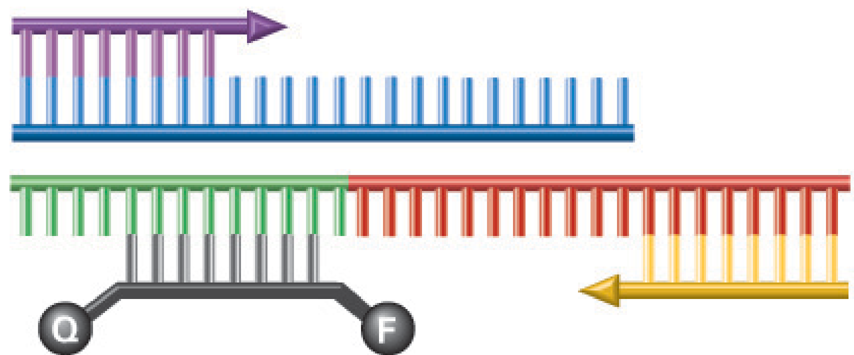



**Figure 5. Diagram of the TaqMan® detection scheme.**

## Conclusion

By utilizing the broad portfolio of Life Technologies assay and sequencing technologies, it has been possible to proceed from a clinical isolate of a pathogenic bacterium to a specific assay within two weeks. This outcome has been made possible by gold-standard Sanger (CE) sequencing with the Applied Biosystems® 3130 Genetic Analyzer for the initial targeted screen, the ultrafast speed and simplicity of the Ion PGM™ sequencer to complete whole-genome sequencing in two geographically distinct sites, the speed and reliability of the Applied Biosystems® 7500 Fast Real-Time PCR System, and the speed and exquisite specificity of custom-designed TaqMan® Assays for ongoing surveillance of the *E. coli* O104:H4 outbreak. Throughout this serious health threat, Life Technologies offered its unwavering commitment to clinical research and public health centers, facilitating rapid identification and full genetic characterization of the organism and development of a reliable, strain-specific screening assay.

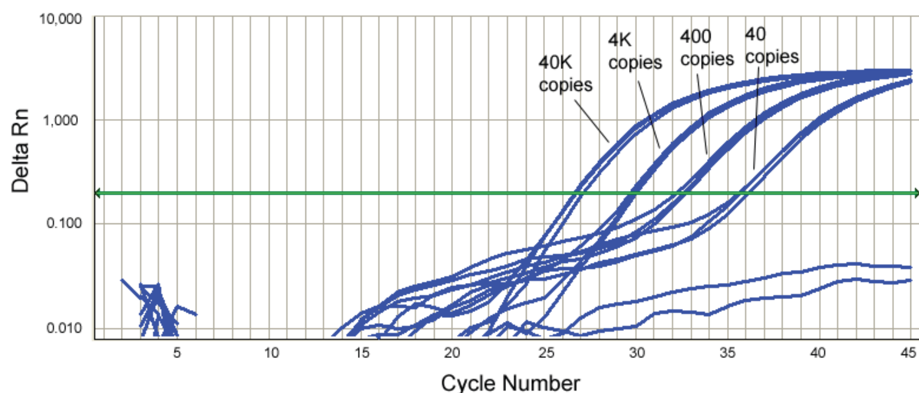



**Figure 6. Newly Designed Custom TaqMan® *E. coli* O104 Assay Detects as Little as 40 copies of *E. coli* O104 DNA.**

## References


1. European Center for Disease Prevention and control Daily Updates June 05 2011
   http://ecdc.europa.eu/en/activities/sciadvice/Lists/ECDC%20Reviews/ECDC_DispForm.aspx?List=512ff74f%2D77d4%2D4ad8%2Db6d6%2Dbf0f23083f30&ID=1100&RootFolder=%2Fen%2Factivities%2Fsciadvice%2FLists%2FECDC%20Reviews
2. Applied Biosystems® Pathogen Detection System
   http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_052984.pdf
3. MLST app note http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_042189.pdf
4. MLST database http://mlst.ucc.ie
5. European Center for Disease Prevention and Control, EHEC infection and antibiotic therapy
   http://ecdc.europa.eu/en/healthtopics/escherichia_coli/clinical_reference_information/Documents/DGI%20position%20paper_EHECantibiotics_English%20version%20plus%20references_20110604.pdf
6. EHEC infection and antibiotic therapy
   http://ecdc.europa.eu/en/healthtopics/escherichia_coli/clinical_reference_information/Documents/DGI%20position%20paper_EHECantibiotics_English%20version%20plus%20references_20110604.pdf
7. Spain seeks compensation for *E. coli* blame
   http://www.bbc.co.uk/news/world-europe-13637130
8. http://ioncommunity.iontorrent.com
9. http://www.iontorrent.com/the-simplest-sequencing-chemistry
10. http://www.ncbi.nlm.nih.gov/nuccore/AFOB00000000
11. http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Link&LinkName=bioproject_sra&from_uid=67657
12. Refseq: NC_011748
13. SAMtools http://samtools.sourceforge.net/
14. MIRA http://sourceforge.net/apps/mediawiki/mira-assembler
15. Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Res.*, 9, 868-877.
16. BGI Sequences Genome of the Deadly *E. Coli* in Germany and Reveals New Super-Toxic Strain
    http://www.bgisequence.com/eu/index.php?cID=194
17. http://www.lifetechnologies.com/news-gallery/press-releases/2011/dna-sequencing-data-reveals-new-hybrid-e.html
18. http://www.ehec.org/pdf/LaborInfo_30052011.pdf
19. http://mummer.sourceforge.net/



**The Life Technologies products discussed are For Research Use Only, and are not intended for any animal or human therapeutic or diagnostic use.**
© 2011 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners. TaqMan is a registered trademark of Roche Molecular Systems, Inc. Printed in the USA. **CO23298 0611**



**Headquarters**
5791 Van Allen Way | Carlsbad, CA 92008 USA | Phone +1.760.603.7200 | Toll Free in the USA 800.955.6288
www.lifetechnologies.com