VTEC strains typing: from traditional methods to NGS

2nd course on bioinformatics tools for Next Generation Sequencing data mining: use of bioinformatics tools for typing pathogenic *E. coli*

ISS, Rome 16-17 June 2016

Subtyping: microbiological and molecular methods for differentiating bacterial species below the species level

Laboratory methods that can differentiate pathogenic microbes (fingerprint) are useful to:

investigate on the transmission of human infectious diseases from different sources

identify community-wide outbreaks

Investigate on the circulation/persistence of specific clones

characteristics for a typing method

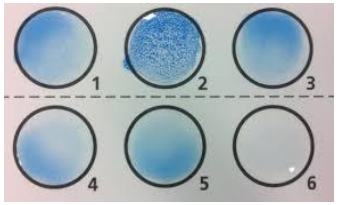
- Discriminatory power
- Reproducibility and repeatability
- Speed
- Low cost
- Ease
- Automation
- Validation and standardization
- Backward compatibility with historical data

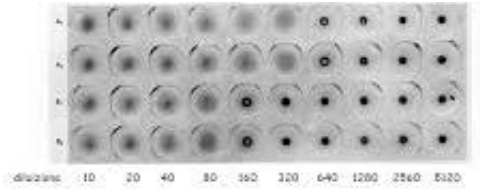
determination of O- and H- Antigens

(Phenotypic and Molecular methods)

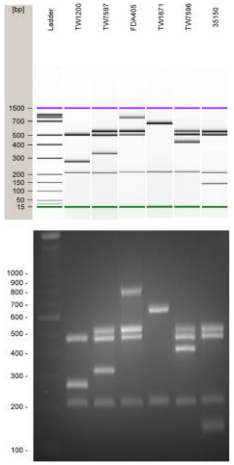
E. coli O antigens: O1 to O188 and H antigens: H1 to H56

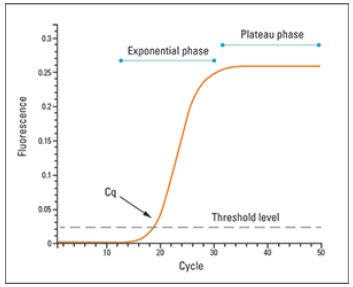
Phenotypic assay: agglutination with antisera





Molecular assay: amplification of O-associated (mainly *wzx* and *wzy*)and *fliC* genes





Serotyping

Advantages

Represents the basis for VTEC strains characterisation

PCR and RT PCR are easy assays

It is of support in epidemiological investigations

Drawbacks

Costly, labour-intensive, cross reaction between different antigens, O non-typeable

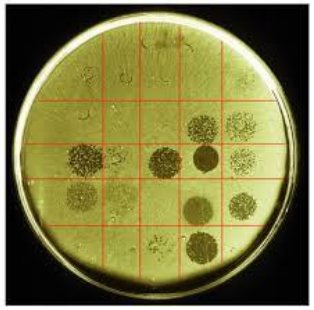
availability of antisera in the lab

lack of PCR tests for the determination of the whole panel

Complete serotyping of E. coli O and H antigens is achieved in a few laboratories

Phage typing

It is a phenotyping method that distinguishes *E. coli* O157 in about 80 PTs based on the susceptibility to the infection of a set of bacteriophages



Advantages

It represented the gold standard basinc characterisation of *E. coli* O157

Drawbacks

Characterises O157 strains only

Performed only in a few reference laboratories

Virulotyping

Mainly obtained by the amplification of virulence genes, to a lesser extent by hybridisation

Verocytotoxin-coding genes: vtx
Colonisation-associated genes: eae, aggR, aaiC, aat, saa, tia, iha, paa
Plasmid-borne genes: e-hly, katP, etpD, toxB
Other toxin-coding genes: subAB, astA, cdt

Advantages

Rapid and easy assays

Availability of a huge amount of already described primer pairs

high throughput Real Time PCR system and microarrays

But also….

Drawbacks

Which targets to characterise the full virulome of STEC??
vtx-phage – LEE – plasmid – OI#122 – OI#57……

Different alleles of the virulence genes exist

vtx, eae, toxB, subAB……

Troubleshooting connected with PCR technique

It may be expensive

Pulsed Field Gel Electrophoresis

Set up 30 years ago

Technique to separate long strands of DNA through an agarose gel by applying a pulsed electric field

PFGE resolves extremely large DNA, raising the upper size limit of DNA separation in agarose from 30-50 kb to well over 10 Mb (10,000 kb)

the most commonly used typing method for outbreak identification, surveillance and investigation for a number of important pathogens, in particular Salmonella, VTEC and Listeria

PFGE currently represents the gold standard for the molecular typing of VTEC and is the method indicated for VTEC in the joint ECDC-EFSA molecular typing data collection

PFGE analysis at a glance…

Bacterial cell suspension

↓ Mix with agarose

Plugs preparation

↓ Lysis and washing

Bacterial DNA in plugs

↓

DNA restriction

↓ plugs are loaded onto agarose gel

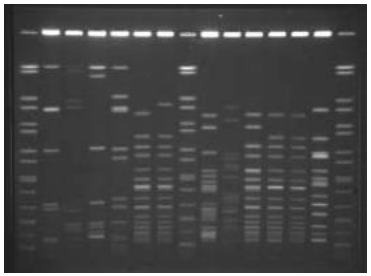
PFGE run

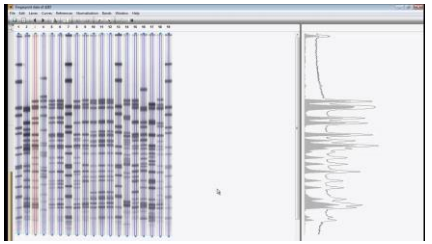
↓ Adequate gel staning/de-staining

Gel image aquisition → Image Analysis

Bacterial DNA is embedded in agarose plugs to avoid shearing and restricted with rare-cutting Endonuclease

Parameters affecting the gel run: buffer, agarose, angle of the field, voltage, time switch, temperature, time





Advantages

Easily applied to different species

Existance of standard operative procedures

Good discriminatory power, useful in outbreak investigations: the use of PFGE allowed an efficient tracing back of the source of infection in a number of cases

Drawbacks

Labor-intensive

Time-consuming

One mutation can yield differences in fragments

Interpretation of results impossible to automate

Multi-locus variable number tandem repeat (VNTR) analysis (MLVA)

Tandem repeats (TRs) are short DNA sequences repeated end-to-end occurring at **specific sites (loci) on the genome**; the number of the reapets at each locus can vary

The thechnique consists in **PCR** amplification of specific loci, with labelled primers, **capillary electrophoresis** in a DNA sequencer, **Bionumerics** calculations of number of repeats per locus. The result consists in a string of numbers

MLVA profile of strain 1: 6-3-6-4-3-5-4-7
MLVA profile of strain 2: 6-3-4-4-2-5-7-7

Advantages

Easy and rapid

Good discriminatory power, supporting the epidemiological investigations

Drawbacks

Availability of DNA sequencer

PulseNet validated protocol for VTEC O157 (Hyytia-Trees et al. 2006)

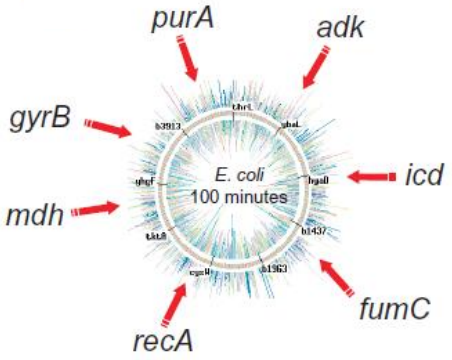
An MLVA scheme was recently published for the typing of VTEC O26 strains (Løbersli et al. 2012)

Brand new: Multiple-locus variable-number tandem repeat analysis for strain discrimination of non-O157 Shiga toxin-producing *Escherichia coli* (Timmons et al. 2016)

Multi-Locus Sequence Typing (MLST)

MLST: analysis of the sequences of internal fragments of seven **house-keeping genes** (genes necessary for organism survival)

- PCR
- Sequencing
- Electropherograms analysis
- Uploading sequences on a webserver to obtain the corresponding alleles and STs



THE UNIVERSITY OF WARWICK

Advantages

Easy to perform

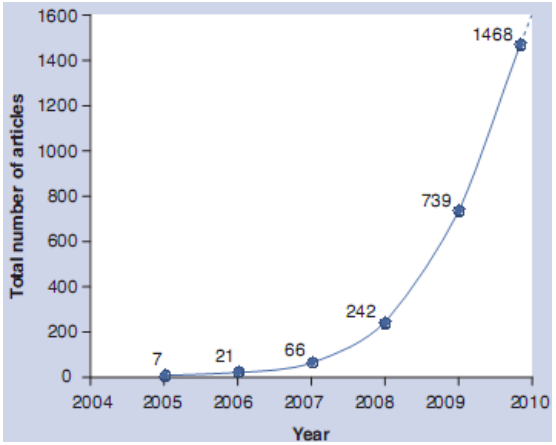
Useful for phylogenetic analyses

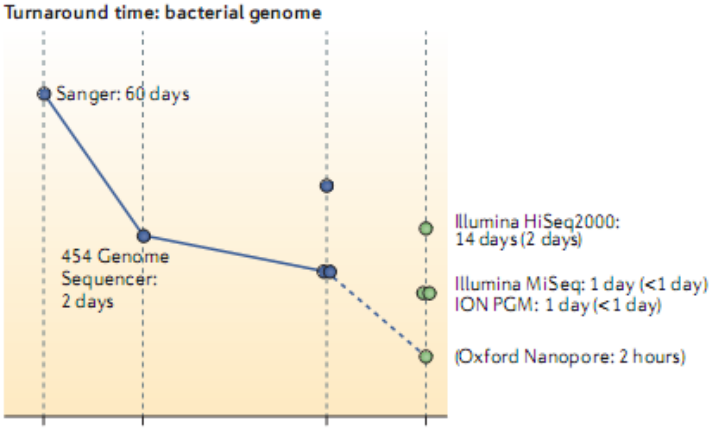
Drawbacks

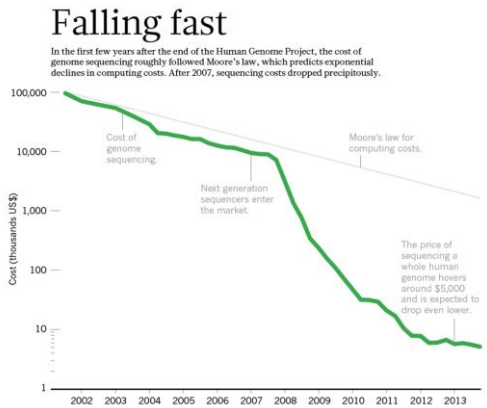
Availability of a sequencer

Not very informative for outbreak investigations

Then comes NGS era







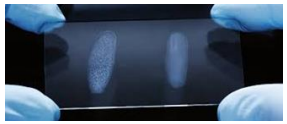
Massive sequence output & low cost per base

Routine WGS of pathogenic *E. coli* and other bacterial species at affordable costs

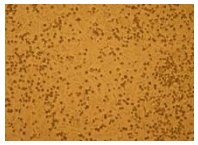
Classical methods and NGS

Isolation of the bacterial pathogen

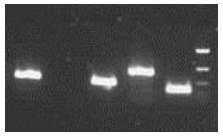
Determine the serotype



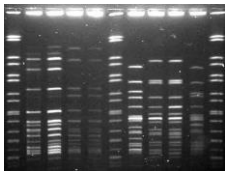
Test suscebtibility to antimicrobial drugs

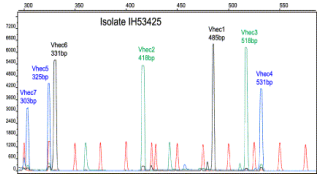


Determine the pathogenic potential



Relate the bacterial pathogen to other strains of the same species



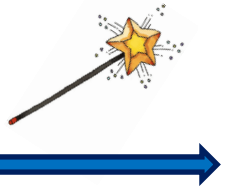


Multistep process

days

---

months



All the results derived from sequencing

Need for huge preliminary work and of intense data analysis

The potential of WGS-based typing is very high

Serotyping

Virulence factors determination

Multiple resistance genes identification

MLVA may be extracted, as well as MLST

But also much more…..

Correlation between strains based on the SNPs in the WGS

Typing based on an extended MLST scheme

Whole genome sequencing typing has the potential to be the new "gold-standard" for pathogen subtyping

BUT

There are challenges that need to be addressed

Data production

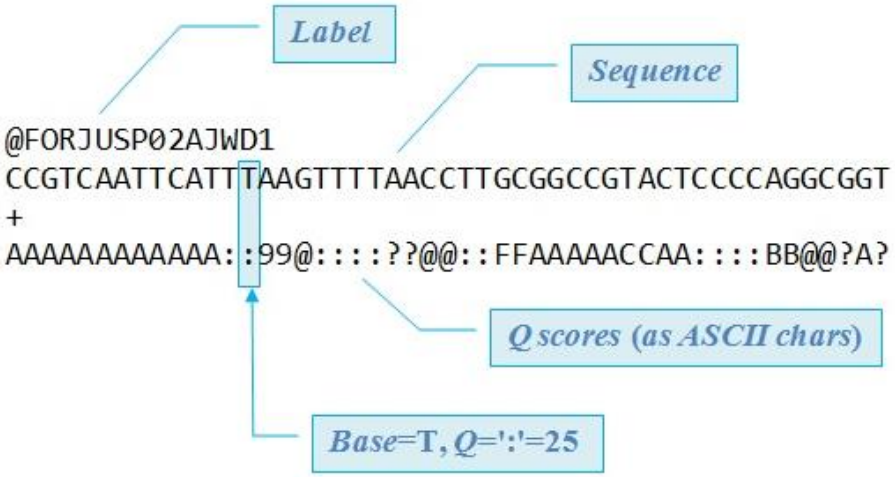
Reference laboratories only actively produce data as of today

Intrinsic quality of the sequence reads at the nucleotidic level

Filtering algorythms to be developed and harmonized

Coverage of sequencing reads (when compared to a reference sequence)

Amount of data produced: storage and transfer



Each .fastq file covering a 5 Mb genome at 30X weights about **300 MB**

Amount of data produced: analysis

Refinement of existing tools for data analysis and development of new ones

Development of analysis pipelines to enhance the assembly, annotation, and interpretation of the data, which will require a coordinated international approach

Computationally intense data analysis
Accessibility of bioinformatic tools via open-source servers

Need for education in bioinformatics

Thank you for your attention!