Molecular typing of VTEC:

from PFGE to NGS-based phylogeny

Valeria Michelacci

10th Annual Workshop of the National Reference Laboratories for *E. coli* in the EU

Rome, November 5th 2015

Molecular typing

VTEC subtyping applications in molecular epidemiology:

-Identification of relationships among microorganisms associated with disease
-Identification of community-wide outbreaks
-Identification of sources of infections
-Establishing geographic origin and distribution of subtypes

Population genetic studies:

-Inference on phylogenetic and evolutionary relationships

HEALTH AND CONSUMERS
Food

2012

http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf.

Vision paper on the development of data bases for molecular testing of foodborne pathogens in view of outbreak preparedness.

1. Background

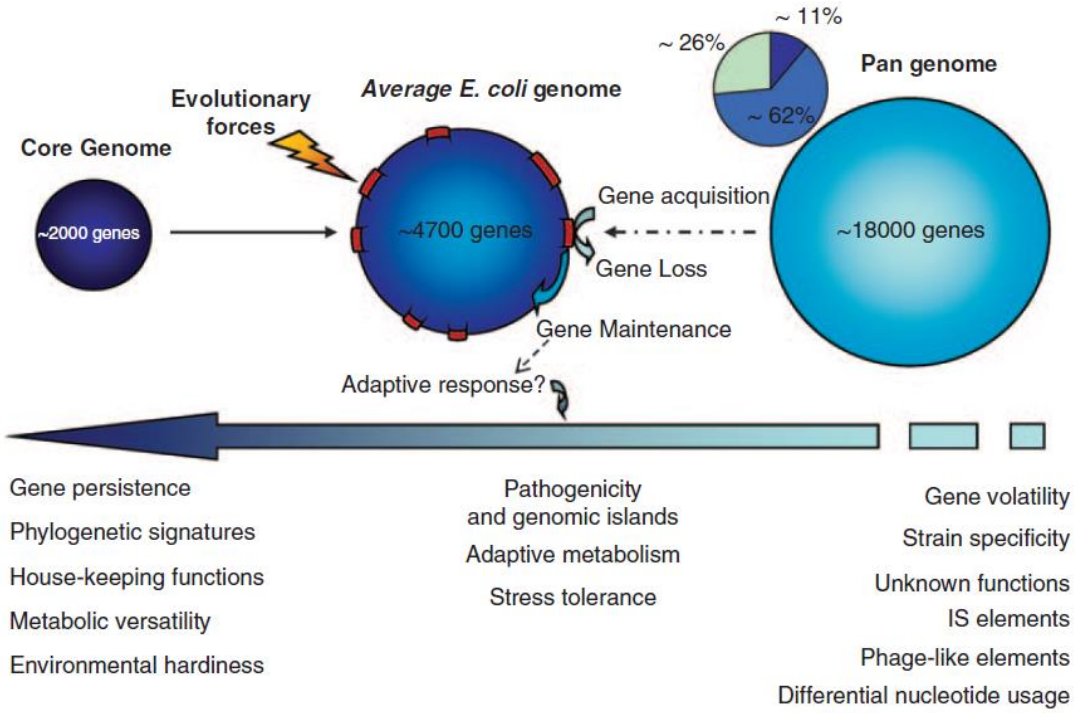
1.1. Molecular testing

Molecular typing or microbial DNA fingerprinting has developed rapidly in recent years. Many typing methods, like PCR techniques and sequencing, have become part of routine strain characterization in many laboratories. Molecular typing provides essential tools for different surveillance purposes, such as monitoring spread of clones and strains, early detection of dispersed (international) outbreaks, and prediction of epidemic potential.

The *E. coli* pangenome

Genomic plasticity

Huge pangenome



Van Elsas J.D. et al., 2011

Pangenome **Whole genome** **Core genome** **Accessory genome**

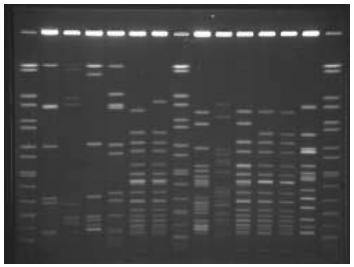
Housekeeping

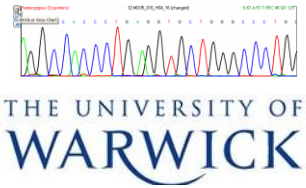
Typing VTEC strains

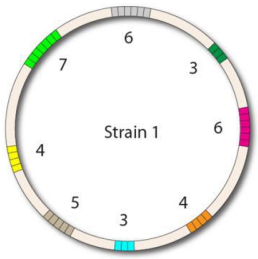
Strain typing takes advantage of the bacterial diversity

- Phenotipic diversity (e.g MLEE, Serotyping, Phage typing)

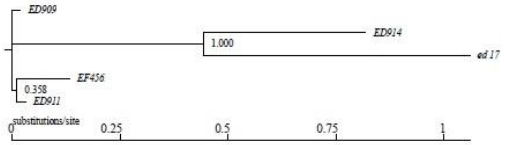
- Genotipic diversity
  - ✓ PFGE
  - ✓ MLST
  - ✓ MLVA
  - ✓ wgSNPs







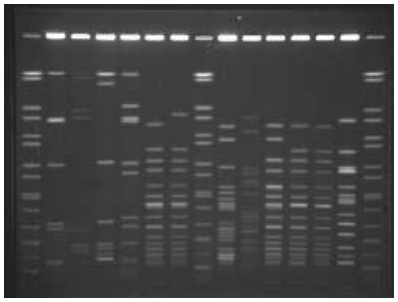
6-3-6-4-3-5-4-7



Pulsed Field Gel Electophoresis (PFGE)

Separation of **large fragments of DNA** (> 50 kb)

The number and size of fragments depend on the position in the chromosome of restriction sites targeted by rare cutting enzymes scattered



PFGE Advantages

- Easily applied to different species
- Existance of standard operative procedures
- Good discriminatory power
- Relatively inexpensive

PFGE Drawbacks

- Labor-intensive
- Relatively slow
- One mutation can yield differences in several fragments
- Interpretation of results impossible to automate

PFGE currently represents the gold standard for the molecular of VTEC

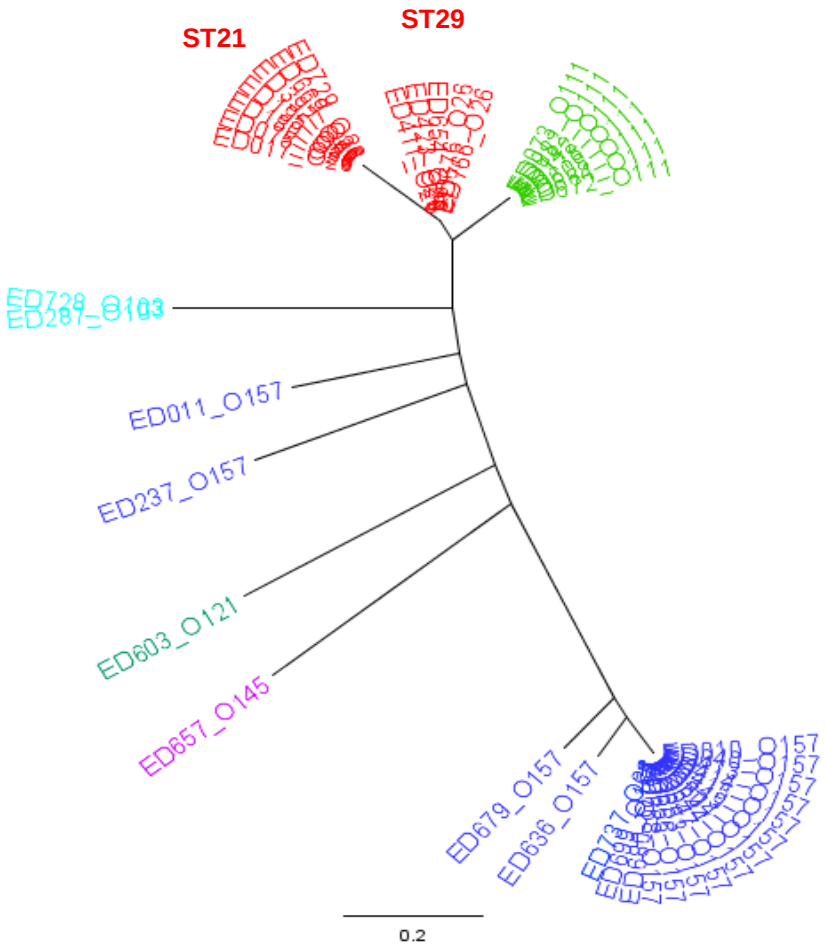
Multi-Locus Sequence Typing (MLST)

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

Two protocols with different sets of loci existing and standardized

THE UNIVERSITY OF WARWICK



MLST in NGS era:

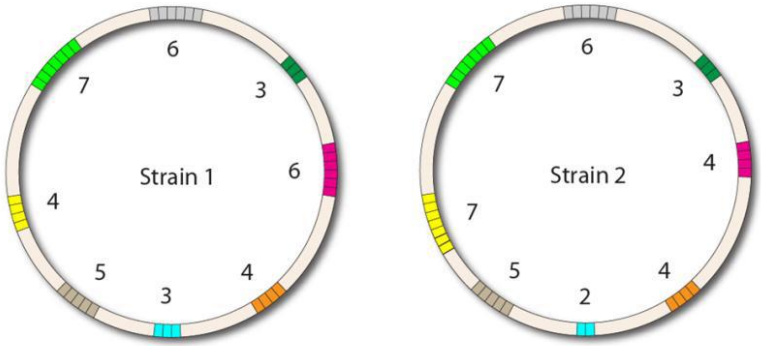
- Extraction from WGS
- NGS of the 7 genes amplified

Good for phylogenetic analyses

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

- **PCR**, **capillary electrophoresis** in a DNA sequencer, **Bionumerics** calculations of number of repeats per locus. The result consists in a string of numbers



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

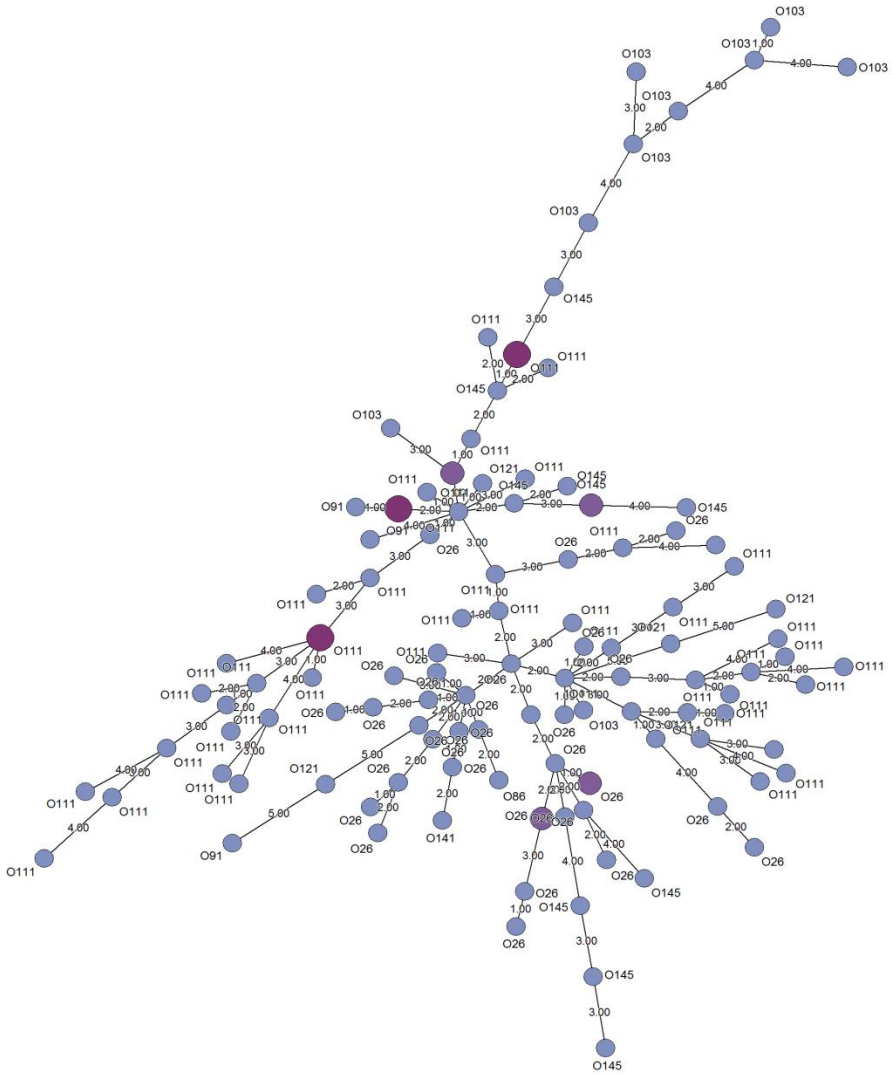
8 loci

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

10 loci

7 from O157 scheme by Lindstedt 2007
1 from Pulsenet
2 additional

MLVA scheme for non-O157: preliminary data



Collaboration with:

10 loci combination of already described schemes and newly designed

Good discriminatory power

Validation needed

Typing through whole genome sequencing

Whole genome Single Nucleotidic Polymorphisms analysis

Reference-based

- **Alignment** to a reference sequence
- Compiling of a **variant call format file** per strain
- Compiling of a **distance matrix**
- **Phylogenetic tree** built on the distance matrix

Reference-free

- Ksnp3 looks for SNPs in central positions of k-mers
- **Different clustering algorithms** available

Tools available for download – possibility to build your own pipeline

Available for download as a tool package operated via command line

CGE webserver hosted by DTU offers easy to use pipelines

Available on ARIES (www.iss.it/site/aries)

Nomenclature for ref-based wgSNP typing: SNP address

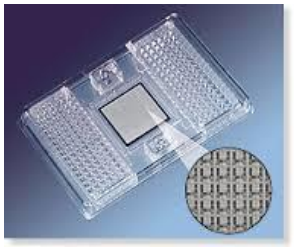
Δ250, Δ100, Δ50, Δ25, Δ10, Δ5, Δ0

e.g.: **25.5.6.48.12.5** and **25.5.6.48.12.9** differ for less than 5 SNPs

Hierarchical single linkage clustering performed on the pairwise SNP difference at various distance thresholds (Dallman T., Microbial Genomics, 2015) Currently published for O157 and apparently also useful for O26

High Resolution Virulence Allelic Profile (HReVAP)

Study on the three major PAIs of STEC: LEE (38 ORFs), OI-122 (12 ORFs) and OI-57 (41 ORFs)

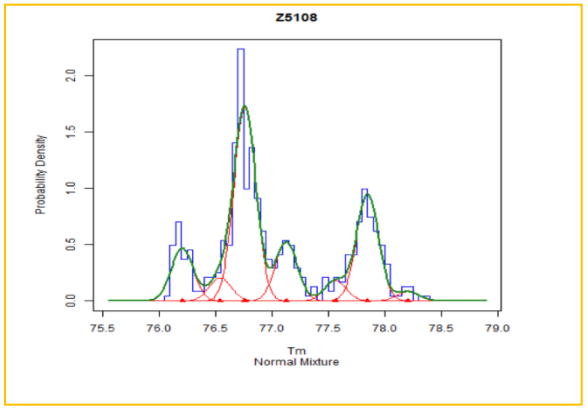


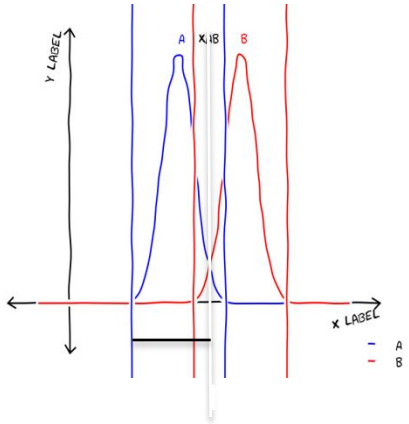
Tm

		ORF1	ORF2	ORF3	ORF4	ORF5	ORF6	ORF7	ORF8	ORF9	ORF10	ORF11	ORF12
O108:H9	DG233/8		77,9		73,7	77,26	78,65	80,78	83,94		76,2	78,04	77,9
O108:H9	DG239/5		78,01		73,78	77,35	78,74	80,87	84,05		76,3	78,13	78,01
O108:H9	DG258/1		78,04		73,81	77,39	78,77	80,9	84,1		76,33	78,18	78,04
O108:H9	DG314/6		78,04		73,81	77,38	78,77	80,89	84,09		76,34	78,17	78,04

Alleles

O108:H9	DG233/8	Z	B	Z	B	C	F	Z	B	B	C	Z	Z
O108:H9	DG239/5	Z	B	Z	B	C	F	Z	B	B	C	Z	Z
O108:H9	DG258/1	Z	B	Z	B	C	F	Z	B	B	B	Z	Z
O108:H9	DG314/6	Z	B	Z	B	C	F	Z	C	B	B	Z	Z





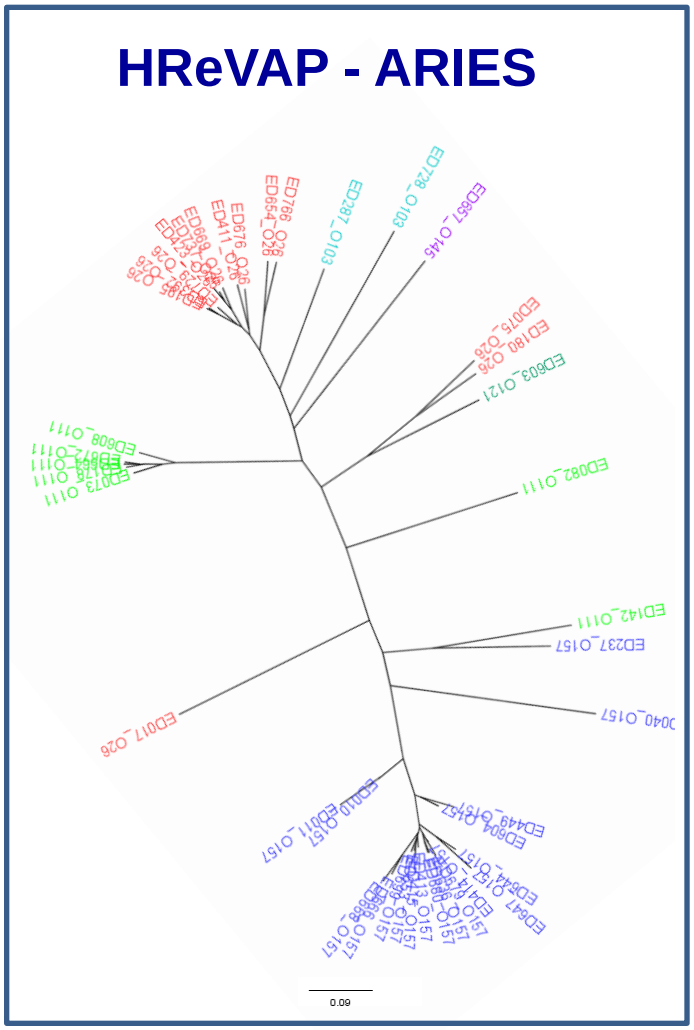
The combination of alleles represents a signature of the tested strain

HReVAP clustering

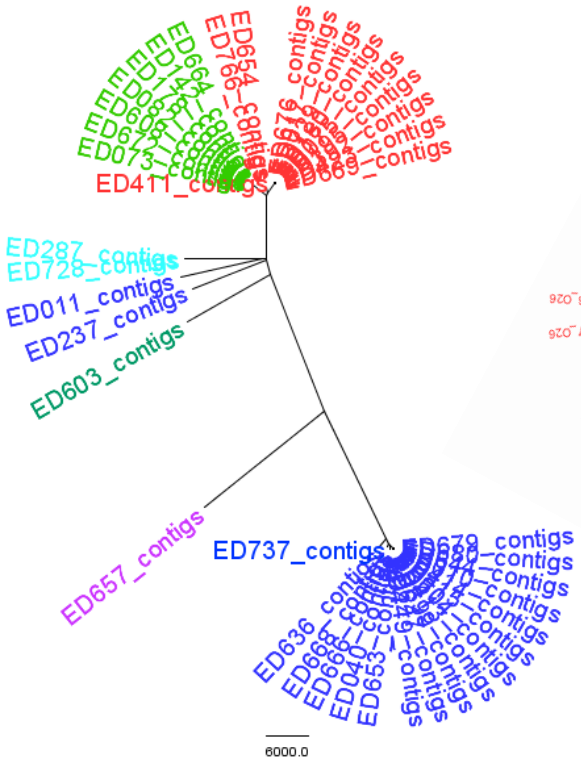
Cluster analyses: HReVAP & SNPs trees

O157 O111 O145 O26 O103 O121

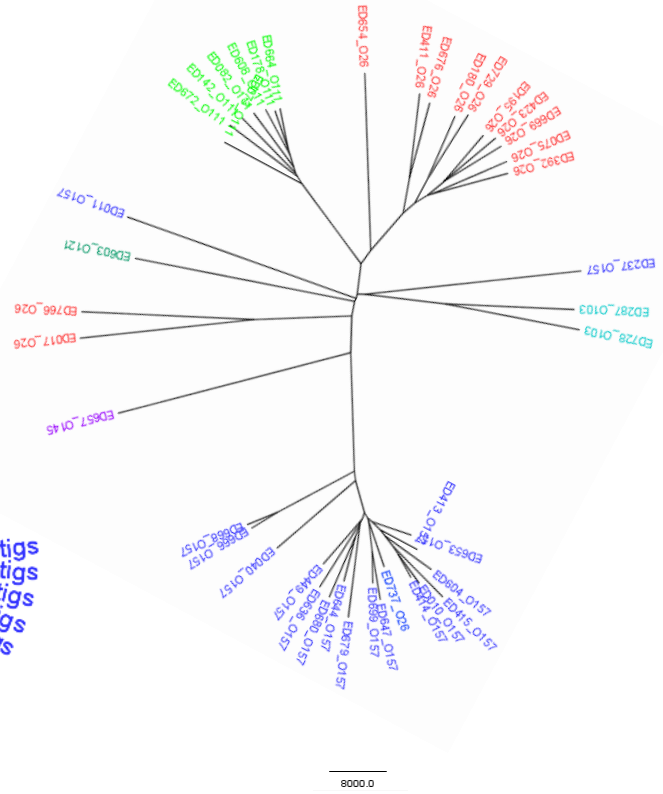
HReVAP - ARIES



Ref based SNPs
Ndtree - DTU



No Ref SNPs
Ksnp3 - ARIES



Conclusions: typing methods for *E. coli* at a glance

	PFGE	MLST	MLVA	wgSNP	HReVAP
Discriminatory level	Strain typing	Phylogenetic analysis	O157 and O26 strain typing Work in progress for the others	Ref-based strain typing pipeline for O157 and O26	Strain typing
Possibility to automate	No	Yes (especially if NGS-based)	Yes	Yes	Yes
Standardization	Yes	Yes	Yes for O157	Not yet	Not yet

Desired characteristics for a typing method

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