Introduction to High Resolution Virulence Allelic Profiling (HReVAP)

Valeria Michelacci   NGS course, June 2016

Istituto Superiore di Sanità, Dip. Sanità Pubblica Veterinaria e Sicurezza Alimentare, Laboratorio Europeo e Nazionale di Riferimento per *E. coli*

The *E. coli* pangenome

Genomic plasticity

Huge pangenome

Van Elsas J.D. et al., 2011
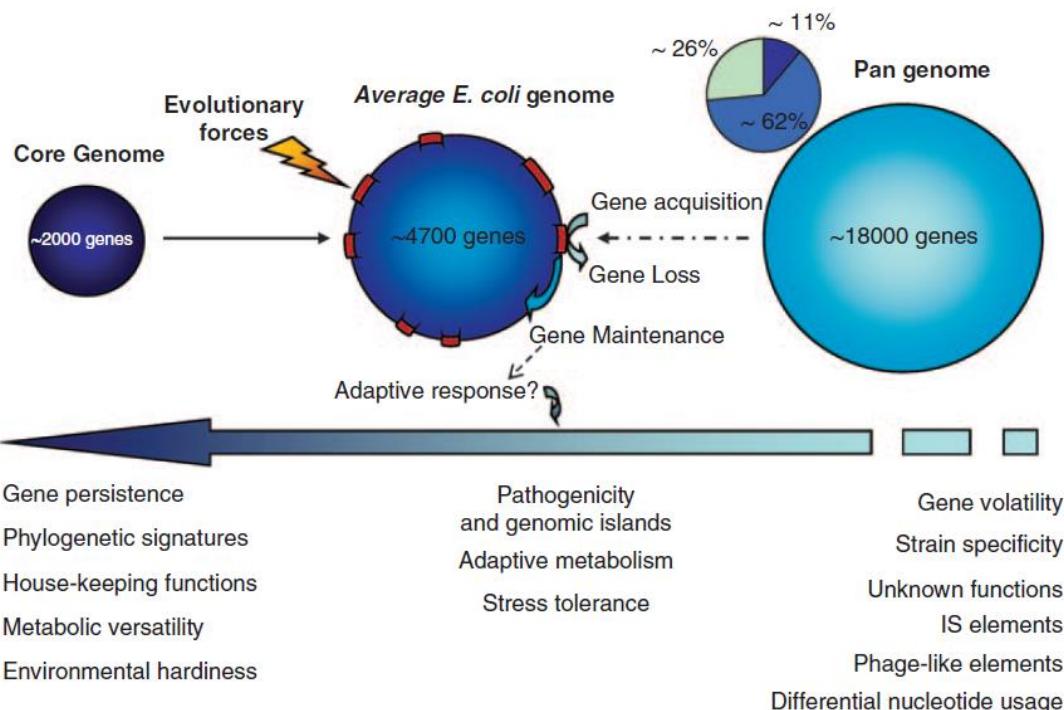
Pangenome   **Whole genome**   **Core genome**   **Accessory**
genome

Housekeeping

STEC accessory genome

- STEC are zoonotic pathogens able to cause severe disease. STEC virulence genes are part of *E. coli* accessory genome

- The existence of allelic variants of STEC virulence genes is a source of variability

stx	*stx1*	*a, c, d*
	stx2	*a, b, c, d, e, f, g*

eae   18 intimin types

subAB $subAB_1$, $subAB_{2-1}$, $subAB_{2-2}$

toxB $toxB_1$, $toxB_2$

Alleles segregate in different STEC groups

The high number of STEC virulence genes and their variability can be exploited for wide scale studies investigating STEC evolution

EU-RL
VTEC

ISTITUTO SVPERIORE DI SANITA

High Resolution Virulence Allelic Profile (HReVAP)

Study on the three major PAIs of STEC

High throughput qPCR: presence and melting temperature analysis
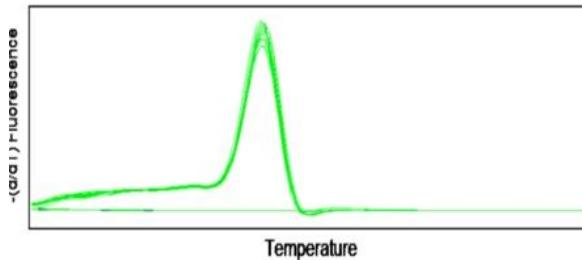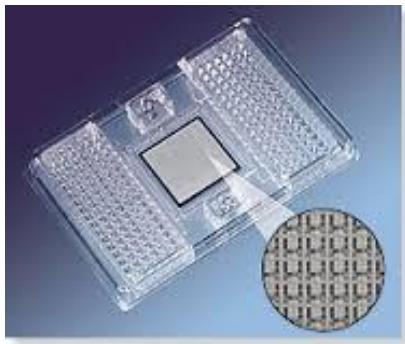
- **LEE**   38/43 ORFs
- **OI-122**   12/16 ORFs
- **OI-57**   41/117 ORFs

TOTAL   **91/172 ORFs tested**

Primers designed on the sequence of EDL933

HReVAP for producing strain signatures of 730 strains: *eae*-positive and *eae*-negative STEC (368 strains) and EPEC (362 strains)

HReVAP: Automatic BIN assigment

|		ORF1	ORF2	ORF3	ORF4	ORF5	ORF6	ORF7	ORF8	ORF9	ORF10	ORF11	ORF12
Tm **Strains**	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 **Strains**	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
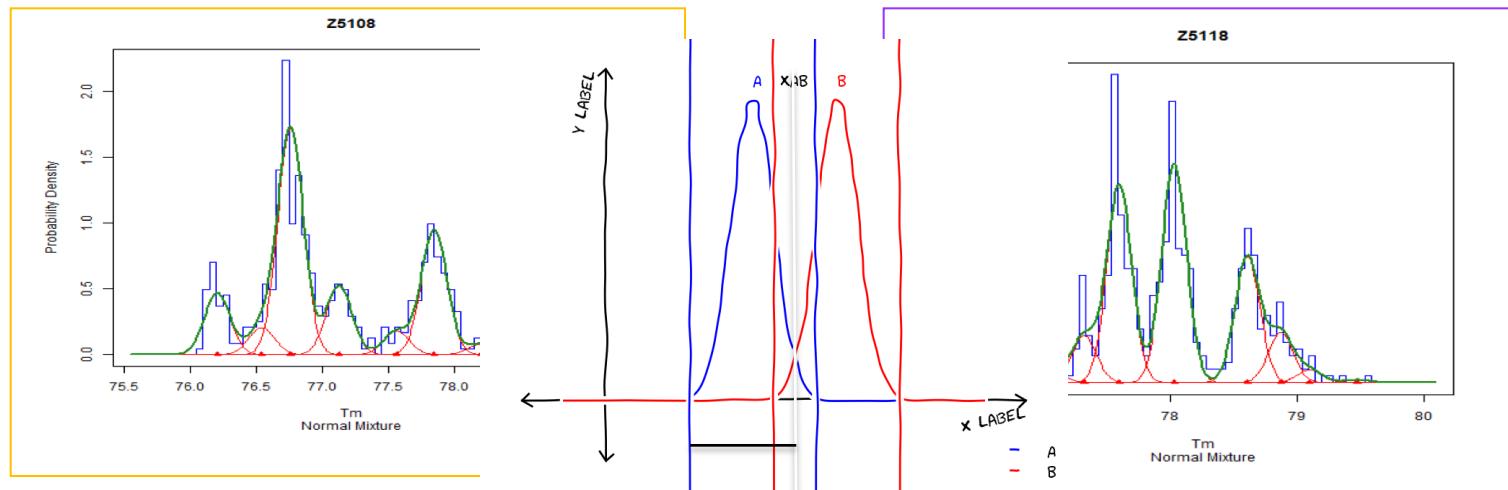
2 to 10 different alleles detected in each of the 91 genes tested (mean = 5.23; tot=476)

The combination of alleles represents a signature of the tested strain

HReVAP clustering

HReVAP: *eae*-positive STEC (three PAIs)

91 ORFs

O157   O111   O145
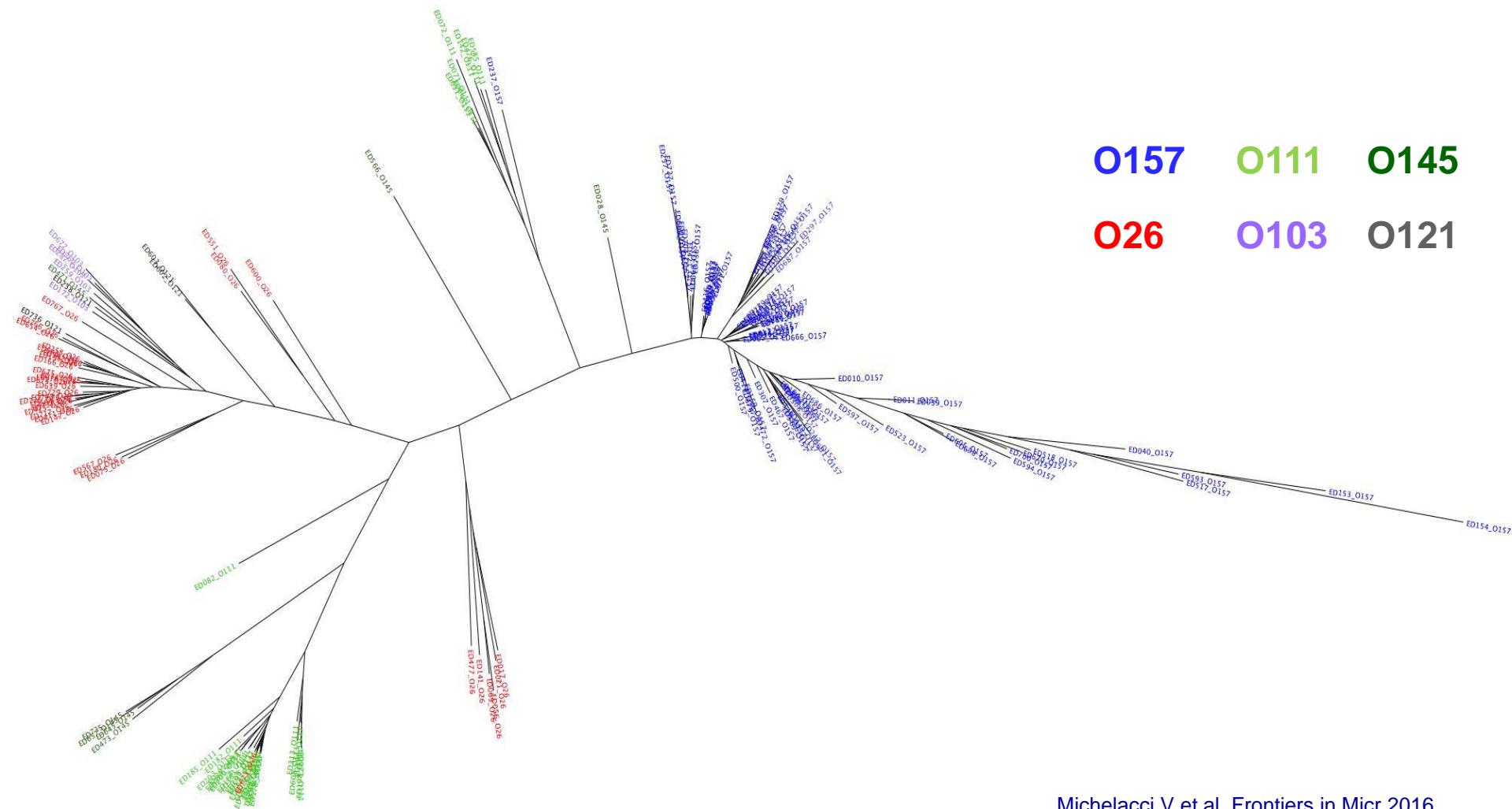
O26   O103   O121

Comparison: HReVAP and NGS

O157 **O111** **O145** **O26** **O103** **O121**

HReVAP - ARIES
Michelacci V et al, Frontiers in Micr 2016
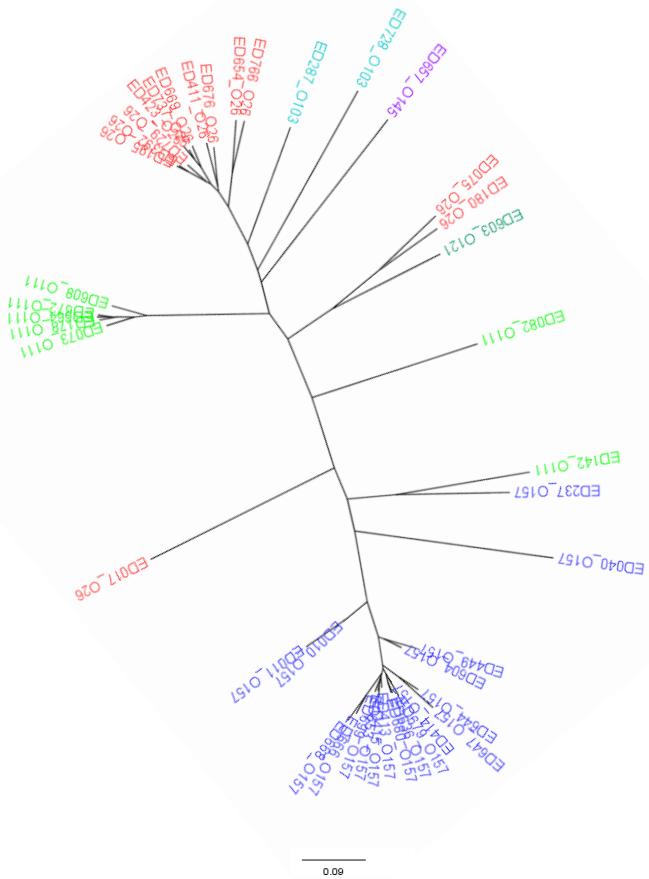
ksnp3
Gardner SN et al, Bioinformatics 2015

Istituto Superiore di Sanità, Dip. Sanità Pubblica Veterinaria e Sicurezza Alimentare, Laboratorio Europeo e Nazionale di Riferimento per *E. coli*
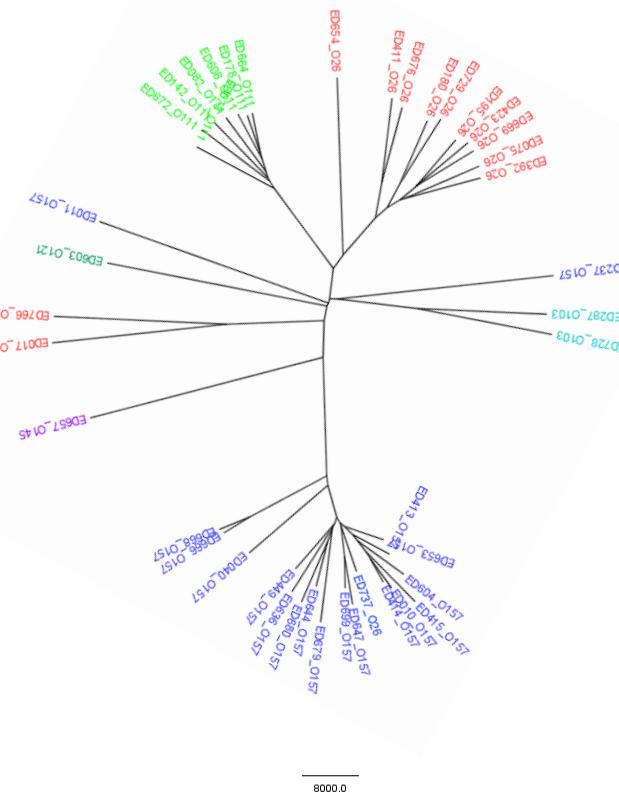
Conclusions

- HReVAP allows following the **evolution of the MGEs** by using subpanels of PAIs for the analysis

- HReVAP identifies sub-populations of STEC even within serogroups.

CONCEPT: Accessory genome is informative for STEC and could be used for typing! Exploit variability for typing

HReVAP tools are automatic and available on ARIES.They can be used to apply the same concept to the study of other pathogens
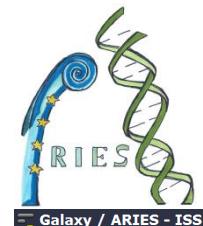
The variability of accessory genome can be studied through NGS

A dedicated MLST scheme can be defined for the accessory genome (agMLST) The new alleles found can be automatically appended in the database.

Possibility to automatically extract the accessory genome database by using tools that extract genomic regions with different GC content

HReVAP – ARIES: how does it work?

Input:
File with dataset of Tm

Create Allele Cluster

Inputs:
- **Own** reference
- File with dataset of Tm (one or more strains)

Allele Cluster

Inputs:
- **System** reference
- File with dataset of Tm (one or more strains)

Allele Cluster

Outputs:
- **Classification (ALLELES)**
- Classes of data
- Reference

Output:
Classification (ALLELES)

Output:
Classification (ALLELES)

(**Workflow: Concatenate termotypes2** of multiple samples if needed)

Input:
File with dataset of alleles

Termotype clustering

When you have a system uploaded!

Acknowledgements

Stefano Morabito

Arnold Knijn

Galaxy / ARIES - ISS Massimiliano Orsini

follow us on twitter
Aries Group @ARIES_GENOMICS

Alfredo Caprioli   Roslen Bondì   Antonia Ricci

Rosangela Tozzoli   Fabio Minelli   Eleonora Mastrorilli

Antonella Maugliani   Paola Chiani

Gaia Scavia   Clarissa Ferreri   Patrick Fach

Federica Gigliucci   Silvia Arancia   Sabine Delannoy