

Next generation sequencing: uno strumento innovativo nella lotta ai microrganismi patogeni

Valeria Michelacci

Roma, 23 Ottobre 2018



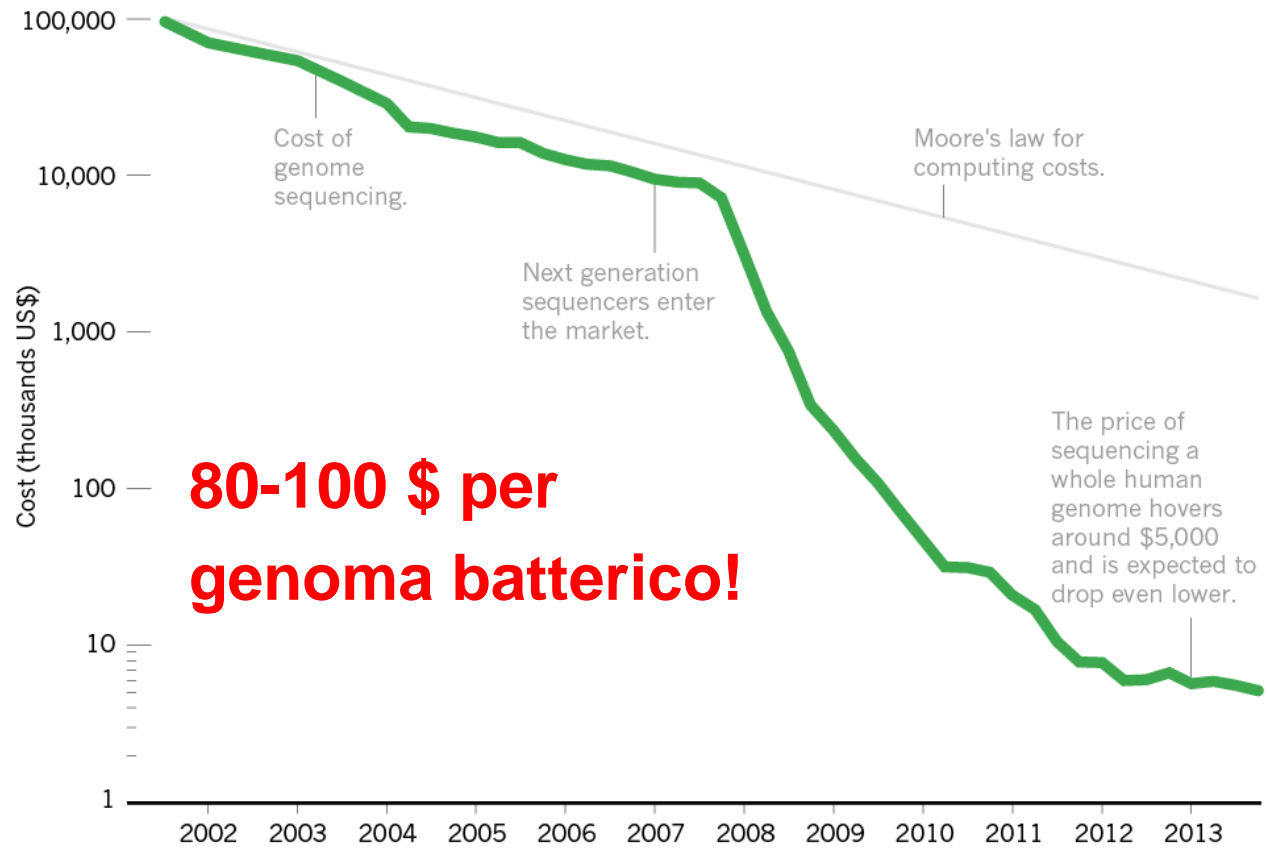
Istituto Superiore di Sanità, Dip. Sanità Pubblica Veterinaria e Sicurezza Alimentare,
Laboratorio Europeo e Nazionale di Riferimento per *E. coli*



Costi associati al sequenziamento di DNA

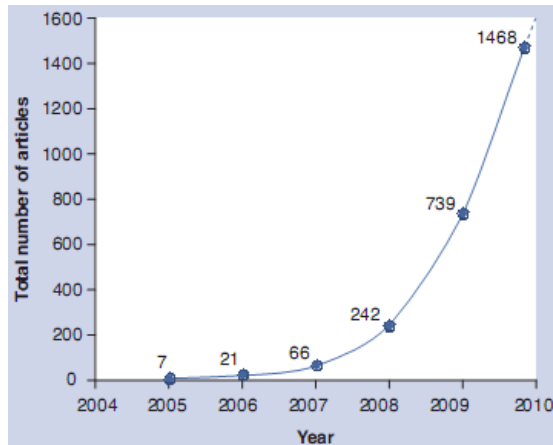
Falling fast

In the first few years after the end of the Human Genome Project, the cost of genome sequencing roughly followed Moore's law, which predicts exponential declines in computing costs. After 2007, sequencing costs dropped precipitously.

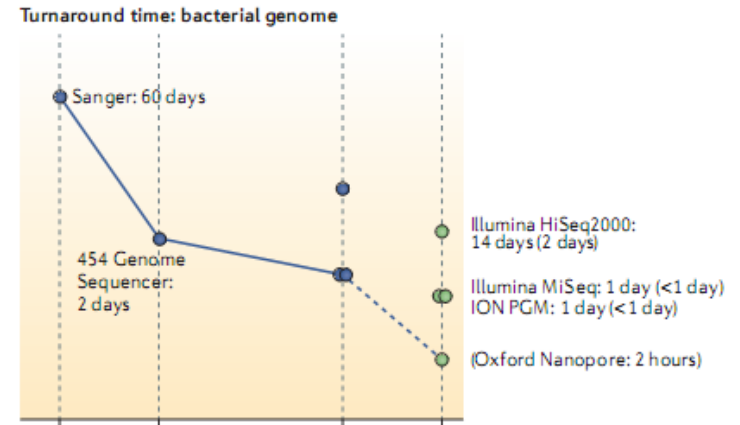


Benefits from NGS

Massive sequence output & low cost/base



Su Z. et al Expert Rev Mol Diagn. 2011 Apr;11(3):333-43.



Didelot X et al. Nat Rev Genet. 2012 Aug 14;13(9):601-12.

NGS in Microbiology

Particularly attractive:

Ability to generate large quantity of **starting material**

Small microbial genomes

Application in microbiology:

Multiple resistance determinants

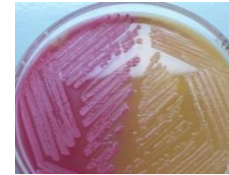
Epidemiological markers

Virulence factors

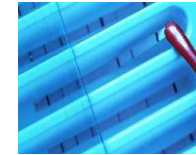
Typing based on the SNPs in the WGS

Metodi classici VS NGS

Isolamento del patogeno

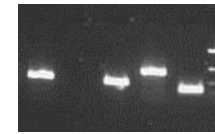


Identificazione della specie

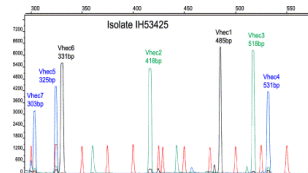
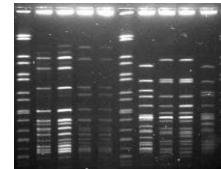


Sensibilità agli antimicrobici

Determinazione del potenziale patogeno



Correlazione con altri ceppi patogeni della sua specie (PFGE, MLST, MLVA...)



Processo multistep

giorni

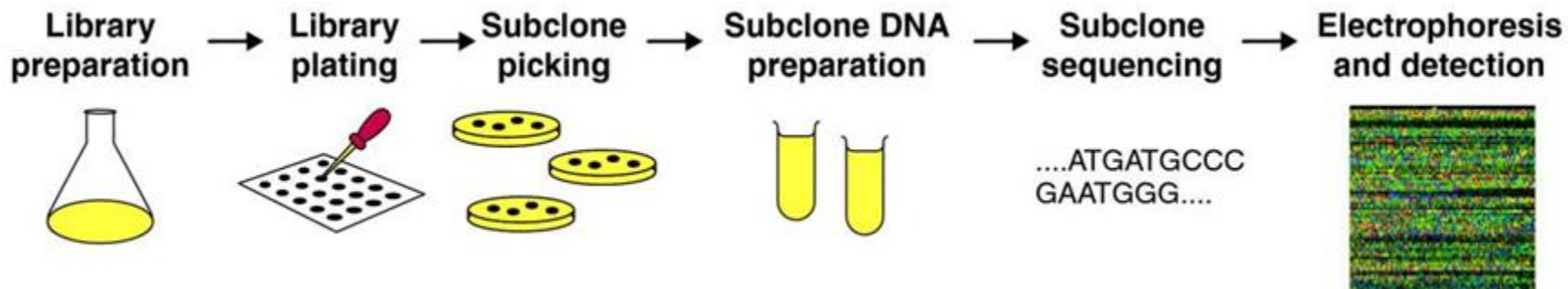
settimane



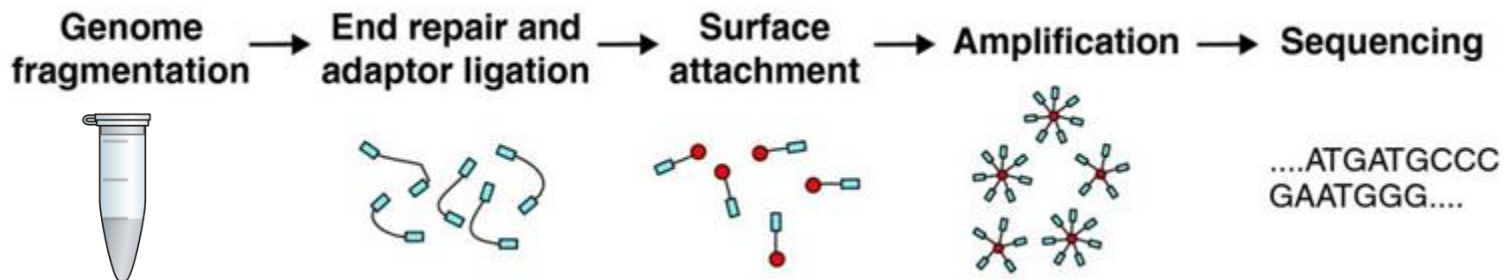
Possibilità di ottenere tutti i risultati con un solo esperimento di sequenziamento!

Sequenziamento convenzionale & NGS

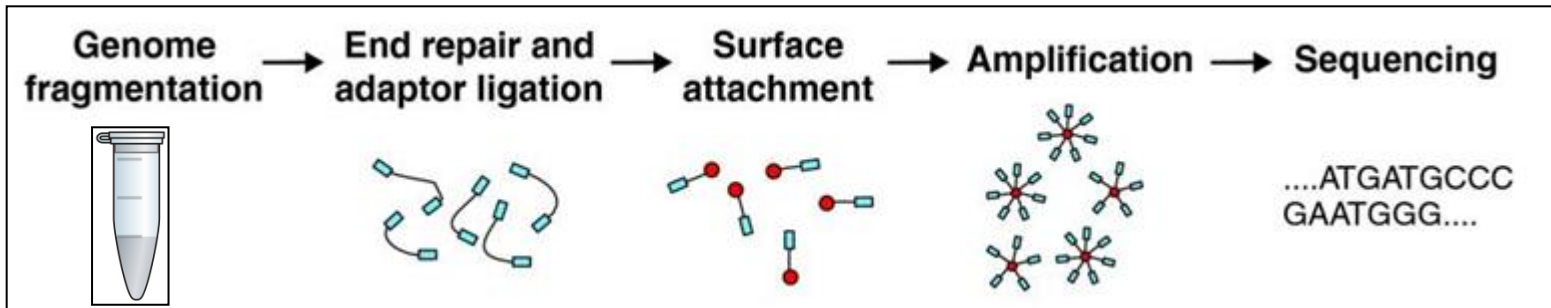
Convenzionale



NGS

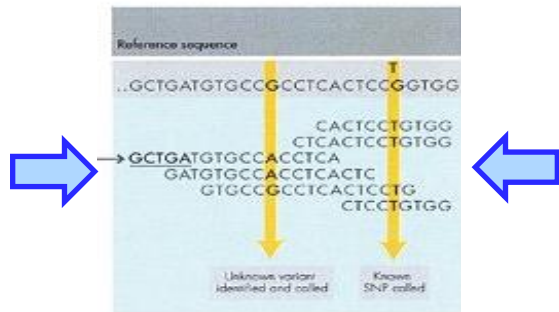


Next generation sequencing



illumina MiSeq

Cattura dell'immagine
Registrazione della fluorescenza

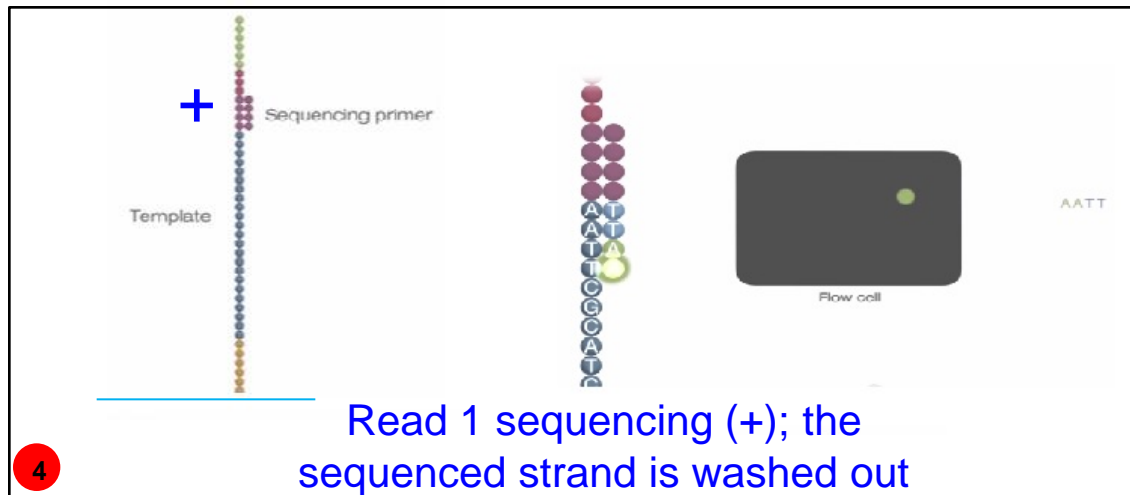
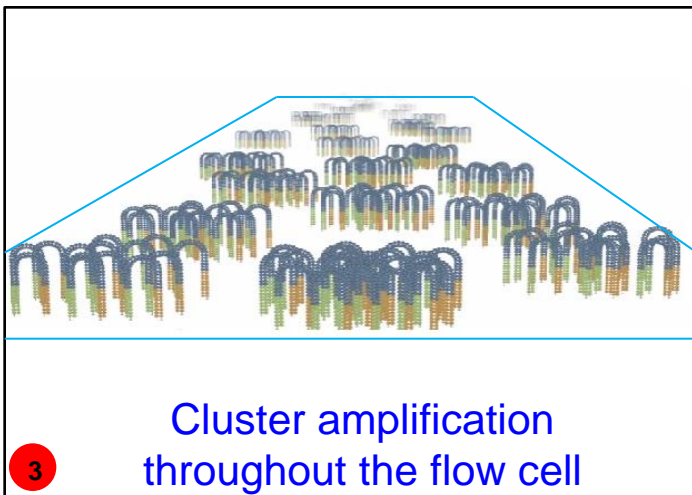
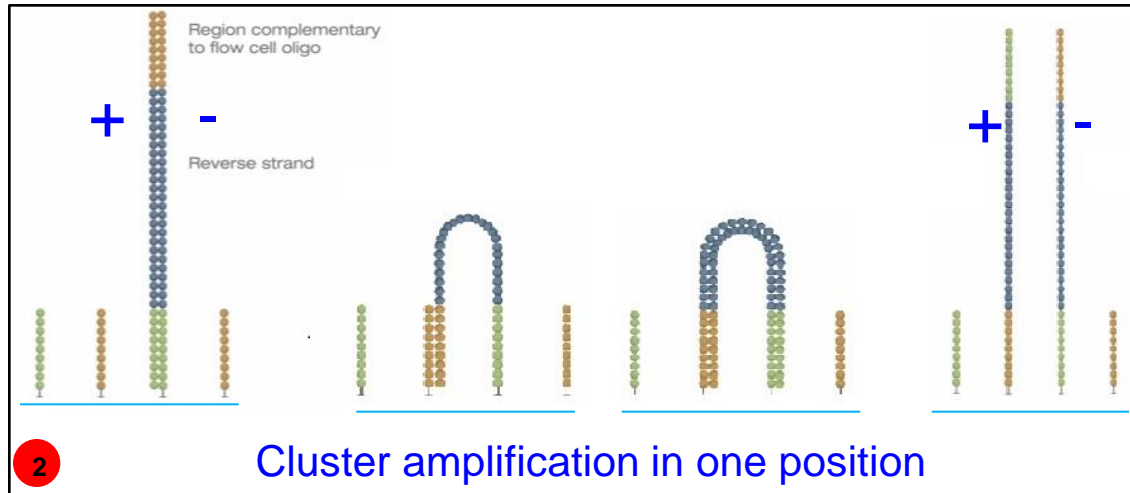
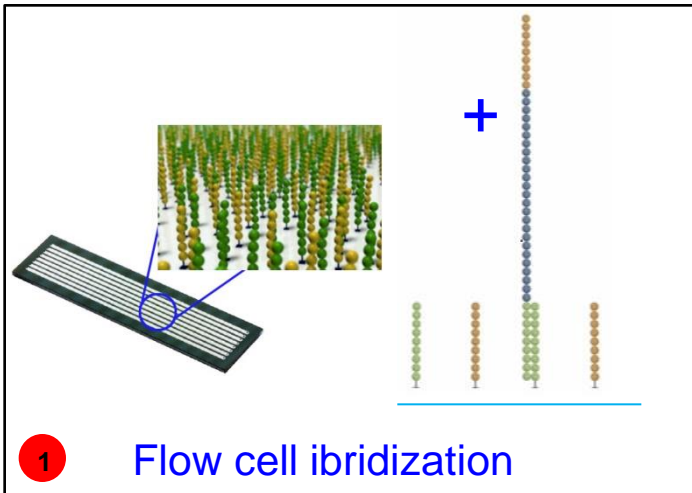


200bp-400bp short reads

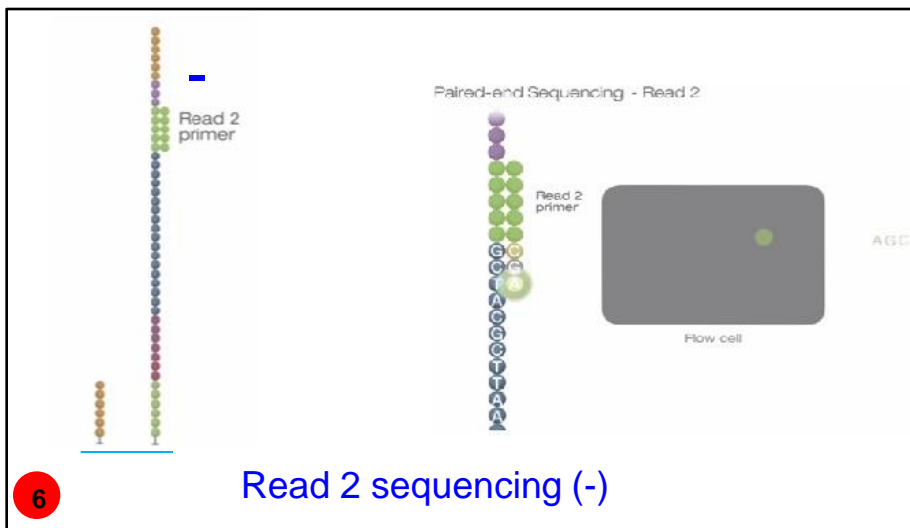
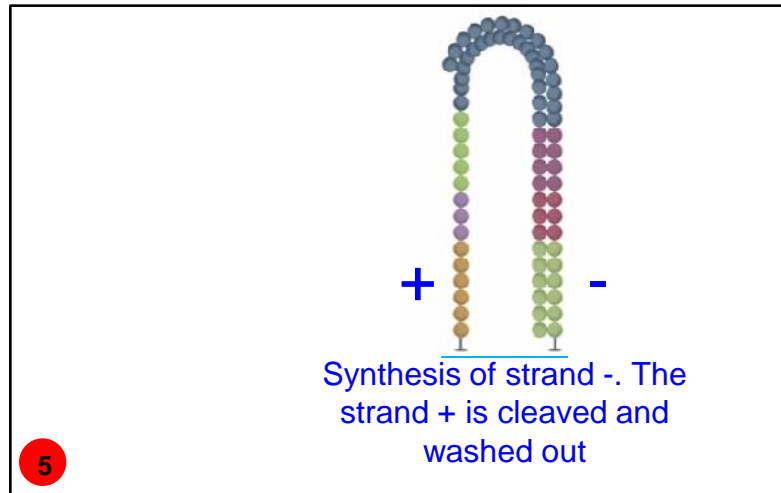
Ion Semiconductor Sequencing Chip

Variazione di pH per incorporazione di nt nel filamento crescente

Illumina Sequencing by Synthesis/1



Illumina Sequencing by Synthesis/2



7

FastQ files compiling

```

strainname_R1.fastq
@read1
AGCTTATCCTCTGCTCACCCCGGGTTAGCGCACTTGATGTATTCACAGC
+
BA1@CC7CBCCC9C8; B2@>C?B@B@B3=9?@B1 : AB7B?B8B?B6B . 7.
@read2
TTGGCGGGGATCTCCAGAAGCATATGGATGTGATCCACACAGCATTCTGC
+
?>?B@<?@, AA7A@C<C?=@@B; +) ?B5* @2=@+=BB, =B6C>AB@B24
@read3
TATGCTCAAGAAGGGGCTGATGAGTTGGTGTGTTTACGATATCACTGCCTC
+
A3AB: B1 : B; 9/0BBBCBB<BB@AA0?BB9: BB<A@BB@7@6@<A@@@<3

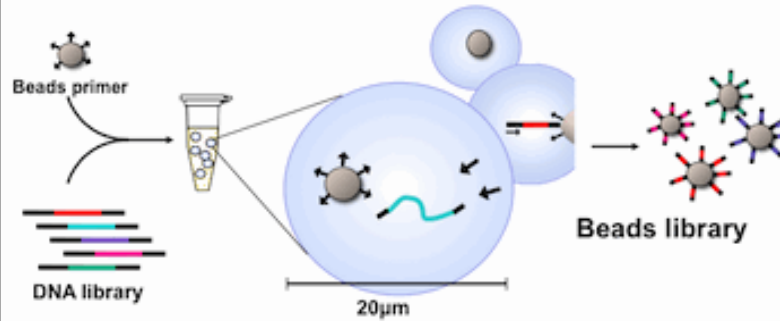
strainname_R2.fastq
@read1
AGCTTATCCTCTGCTCACCCCGGGTTAGCGCACTTGATGTATTCACAGC
+
BA1@CC7CBCCC9C8; B2@>C?B@B@B3=9?@B1 : AB7B?B8B?B6B . 7.
@read2
TTGGCGGGGATCTCCAGAAGCATATGGATGTGATCCACACAGCATTCTGC
+
?>?B@<?@, AA7A@C<C?=@@B; +) ?B5* @2=@+=BB, =B6C>AB@B24
@read3
TATGCTCAAGAAGGGGCTGATGAGTTGGTGTGTTTACGATATCACTGCCTC
+
A3AB: B1 : B; 9/0BBBCBB<BB@AA0?BB9: BB<A@BB@7@6@<A@@@<3
    
```


Ion Torrent semiconductor sequencing/1



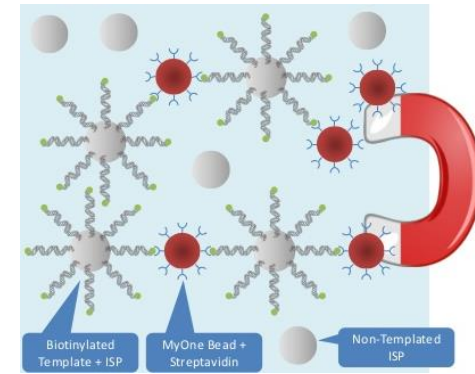
Shearing and adapter ligation

1



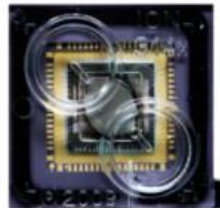
Equimolar emulsion PCR

2

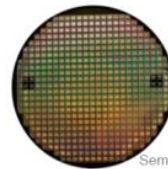


Magnetic enrichment step for loaded spheres

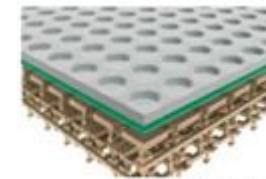
3



Chip
Semiconductor Packag



Wafer
Semiconductor Manufacturing

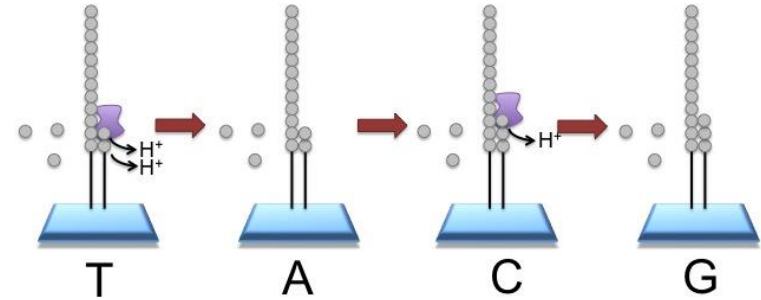
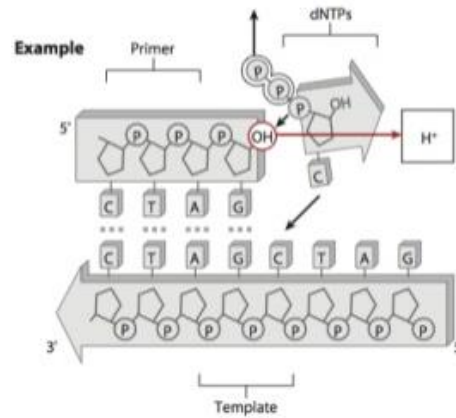
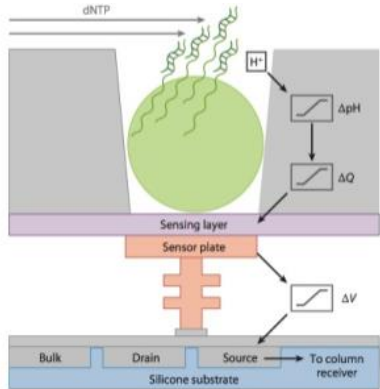


Millions of Sensors

Chip loading

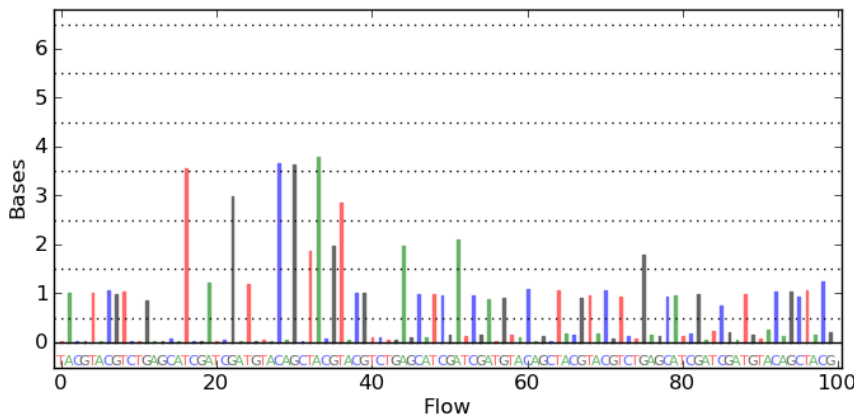
4

Ion Torrent semiconductor sequencing/2



5

dNTPs flow and pH monitoring

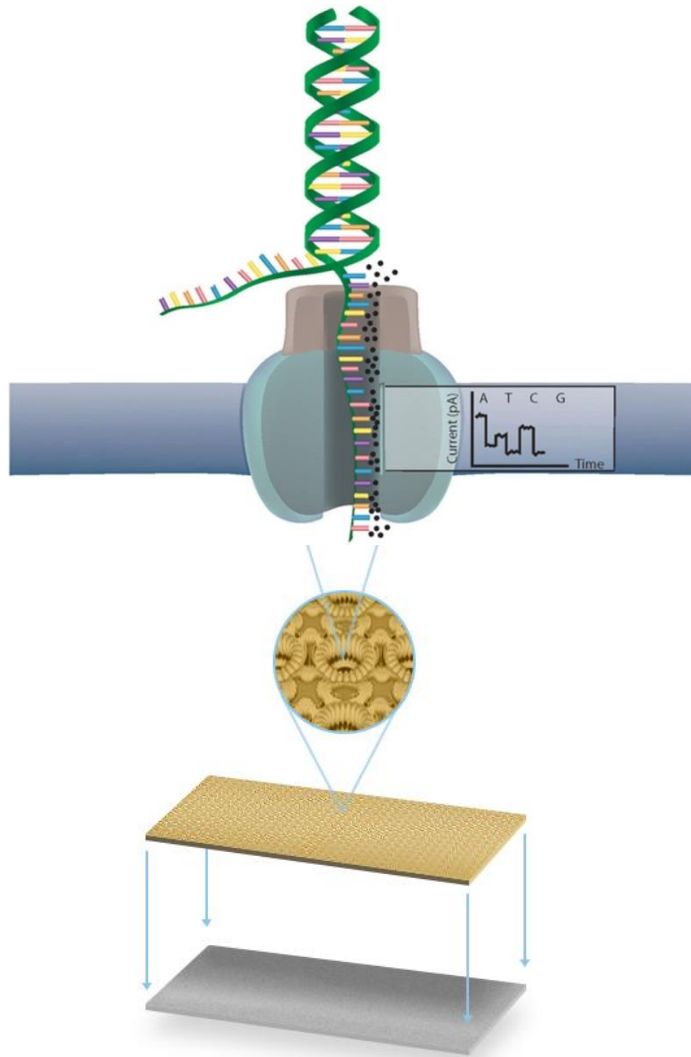


```
@read1
AGCTTATCCTCTGCTCACCCCGGGTTAGCGCACTTGATGTATTACAGC
+
BA1@CC7CBCCC9C8; B2@>C?B@B@B3=9?@B1 : AB7B?B8B?B6B . 7.
@read2
TTGGGCGGGATCTCCAGAAGCATATGGATGTGATCCACACAGCATTCTGC
+
?>?B@)<?@, AA7A@C<C?=@@B; +) ?B5*@2=@+=BB, =B6C>AB@B24
@read3
TATGCTCAAGAAGGGGCTGATGAGTTGGTGTTTTACGATATCACTGCCTC
+
A3AB: B1 : B; 9/0BBBCBB=BB@AA0?BB9: BB<A@BB@7@6@<A@@@<3
```

6

Flow diagram interpretation and fastQ file compiling

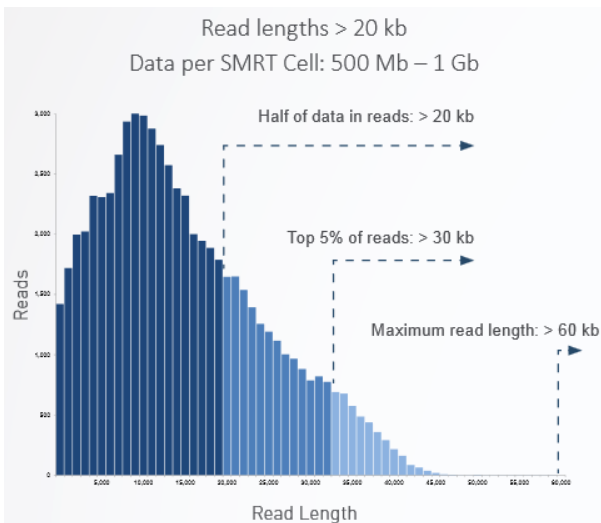
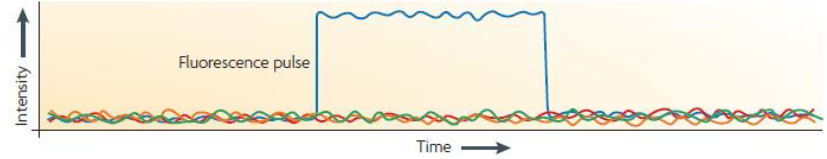
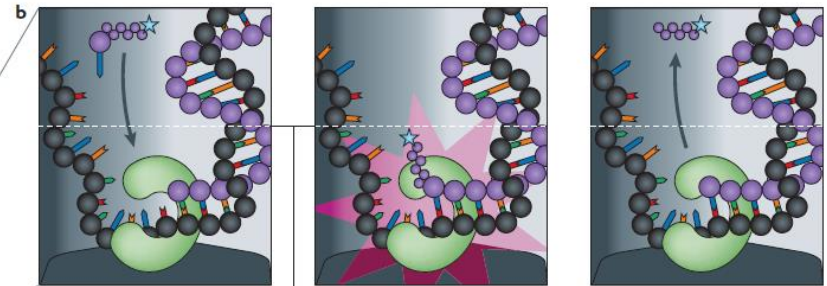
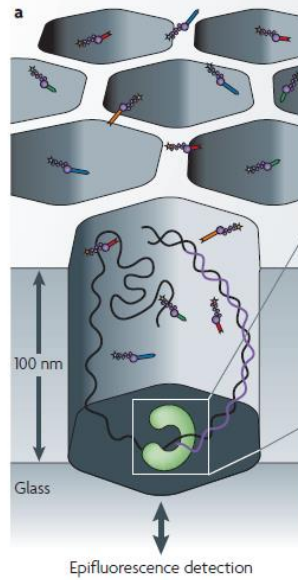
MinION - Oxford Nanopore Technologies



Pacific Biosciences



Pacific Biosciences — Real-time sequencing



Based on data from a 20 kb size-selected human library using a 4-hour movie with P6-C4 chemistry, analyzed with SMRT Analysis v2.3. Each SMRT Cell generates ~55,000 reads.



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



NGS equipments adoption

Next Generation Genomics: World Map of High-throughput Sequencers

Show all platforms 454 HiSeq HiSeq X Ten Illumina GA2 Ion Torrent MiSeq MinION NextSeq PacBio Polonator Proton SOLID Service Provider



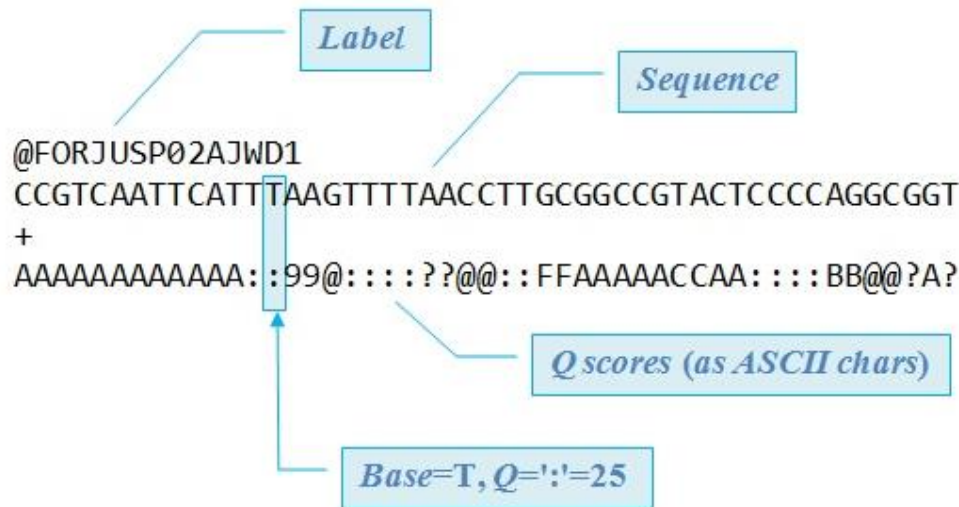
Source: Omicsmap.com June, 17, 2015



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



.fastq files



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

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Phred quality score

$$Q = -10 \log_{10} P$$

from 0 to 93 using ASCII characters 33 to 126

.fastq files

@

```
@X1L6C:01561:00672
AAATATCACCAAATAAAAAACGCCTTAGTAAGTATTTTTCAGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTG
GATTAATAAAAAAGAGTGTCTGATAGCAGCTTCTGAACTGGTTACCTGCCGTGAGTAAATTAATAATTTATTGACTTAGGTCAC
TAAATACTTTAACCAATATAGGCATAGCGCACAGACAGATAAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCA
CCACCATTACCACCACCATCACCATTACCACAGGTAACGGTGC GGCTGACGCGTACAGGAAACACAGAAAAAAGCCCGCA
CCTGACAGTGCGGGCTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCG
```

@

```
+
CC:9:;FBC<CD7:88888(>><C<CCCC<CCBBAAB/A@A8888,;<@;AABBB=?;B98992:B<
CGBBCGDCC?>BCC;BB<ADEEED*CCCAACCCBCABBDDDB>B?>A;999;@8=>199A7>9:;CBCH:B:>>>)999)
77037;<7==5=@@BBCC:C@BBB9B<E<D9>>><<6ADCBCBAABBB@@@DDCBA@==+.//?B<?>AEB:;6;DCD>
C;;;-:9:BC<BBCCC9?><AA;AG<CB>GD@B;;;A<AE;AA<B?>@9@C<BB<?>?BB;BBBAAAA:::BAB099/9>
@=====(<<?)99997>>CCEBA>>=>2373333&3:99-33(3--717--43606704/47761
```

@

```
@X1L6C:01104:03031
AGAAGCTGCTATCAGACACTTTTTTAAATCCACACAGAGACATATTGCCCGTTGCAAGTACAGAAATGAAAAGCTGAAAAATA
CTTACTAAGGCGTTTTTTTATTGGTGATATTTTTTCAATATCATGACAGCAAACGGTGCAACATTGCCGTGTCTCGTTGCTC
TAAAAGCCCCAGGCG
```

@

```
+
@AC=BCCC?>?B?@<CBB@?>>>>>?>8?>>DAABEBCBABCACAA:@@>+9:8>;<///.
98283988*44449;;9/88:~29:>>5;78333333&399298:6/./DCDDCC';>:ACBDAABB?>9:;+<
1444@:~77-3<03368:8755888;;9833)3777'--'--
@X1L6C:03659:02717
```

@

```
GCTTCTGAACTGGTTACCTGCCGTGAGTAAATAAAAATTTATTGACTTAGGTCACTAACTTTAACCAATATAGGCATA
GCGCACAGACAGATAAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCAACATTACCACCACCATCACCATTA
CCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGACCCGCCACTGACCAGTGCG
```

```
+
??>9?BB@<CAA;A8@?>?@5:;BCCCEC;C=CCC8CEJ8DE;AACF>CC?DDCCCB:BB@?>?>9?;B=B=CAA@?;>BCG
CCCCCBABBBBCCDDAA2:4;@?>?>CAB@AAA9@AB?>C;;;C;CDCC>ECCAA<AC<CB>DC<AB=CD=C9:;A4:;>
CC;@@@A?>CI@DDAFKDDD:A@CBDC:;:99199+8;4746@CA?)<444/3:4934333-3888//
@X1L6C:02011:02071
```

```
TTAAATTTTATTGACTTAGGTCACTAAACTTTAACCAATATAGGCATAGCGCACAGACAGATAAAAAATTACAGAGTACA
CAACATCCATGAAACGCATTAGCACCAACATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGTGACGCGTACAG
GAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTCGACCAAAGTAACG
+
=@>>>19;;;7=CCDADC;?:::;5;==4>273:<@BBCF=CDH;@;MMFEED@>>>:::~5/55<
;:@;:;BC=BCBB<B@@@D<@B;3:;9@<BB=BD=AC;@B;?>3:;CAC=CD;;;=BBAB>CC;AA;BAAA9AD@>>
>>>955>4?949998555555&4<>2:;661499888...88/56666666$;6/.5:8(..+'++
@X1L6C:01333:03005
GCAATGCCAGGCAGGGCATGTACAGCTACGTACGTCTGAGCATCGATCGATGTACAGCTACGTACGTCTGAGCATCGATCGA
TGTACAGCTACGTACGTCTGAGCATCGATCGATGTACAGCTACG
```

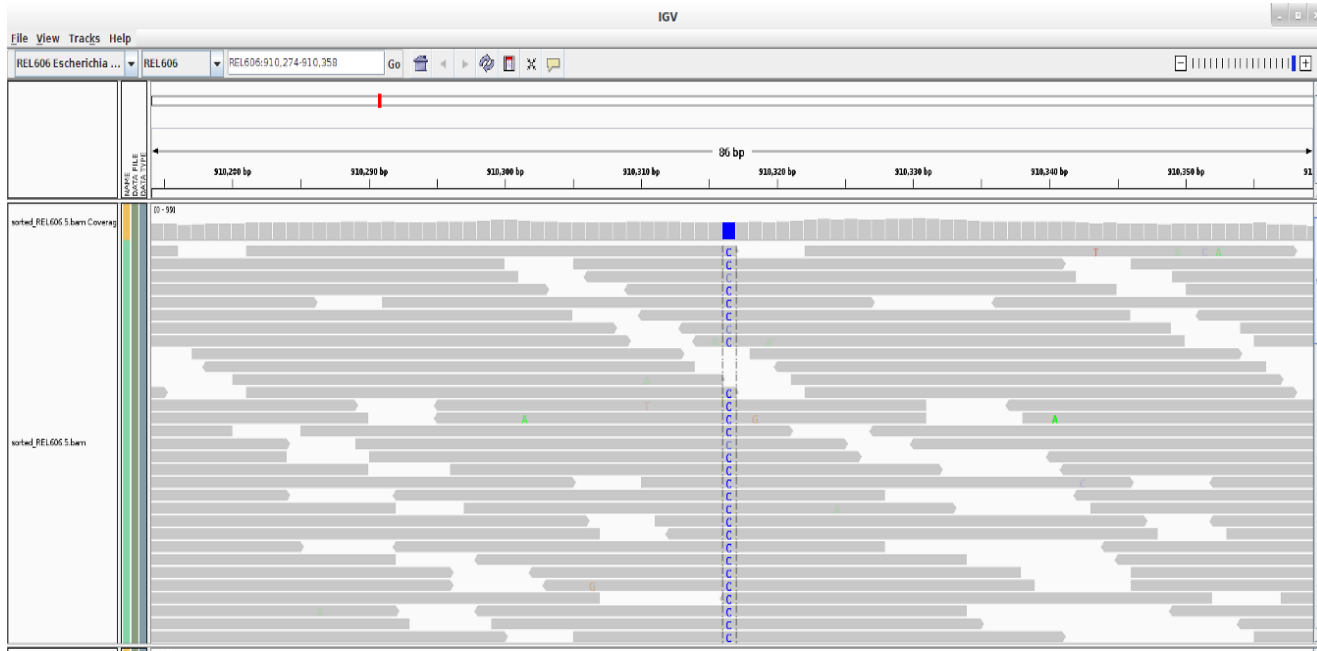
```
+
555/55/(//(/(/(/8:9:<=>><?@:98A??676<;;@:5555555554444;=4443333;383338<68>
68=333111831111111111113933644588?==<76992---2+++0/
```

...and so on



Coverage

Reads mapped on a reference genome



Ref seq

COVERAGE

Mapped reads

Throughput of benchtop sequencers

	Read length	Total time	Total reads	Output
Illumina MiSeq	2 x 300bp	~56 hrs	44-50 M	13.2-15 Gb
Ion Torrent PGM	400bp	~19 hrs	4-5.5 M	1.6-2.2 Gb

e.g. *E. coli* multiplex genomes sequencing

Average length 5 Mb

150 Mb/strain

Desired coverage 30X



Assembly

Short sequencing reads

.fastq file

```
@HWI-ST700693:238:B0224ACXX:1:1101:1218:1982
NACACTTGCTTTGGTGACAGCGGGGCATCCTCAAGC
+
#1=DDDDDHAF?GEFGIIIIIIIIIIIIIIIIIFI
@HWI-ST700693:238:B0224ACXX:1:1101:1161:1986
NGATTTTGACCTCTCCAGTTTCCTCTTAACACTTTC
+
#1:BDFFFGHHGJJJJIJHIJJJJJJJJJJJJJJ
@HWI-ST700693:238:B0224ACXX:1:1101:1193:1989
NTATCCAGCCTGCGGTGCTACTTGGTGAAGAGGAT
+
#1=DDFFFHGHGJJJFGHJJJJIEGECHDFHCC?
@HWI-ST700693:238:B0224ACXX:1:1101:1440:1981
NTCAAGAATCCAAGTGGGGCCAGCATAATGTACGCT
+
#1=DDFFFHGHDFDAEGIIIFGIICGGHGBPFGEFDHI
```



Partially assembled genome (contigs)

.fasta file

```
>NODE 1 length 449 cov 4.835189
ATCTTT CGCGCCTTCCAGCTCCAGCCATT CGGAACCGTT CGCCAGAAAACGGGCGTAATC
GGGT AAGACAT AGCGCGGTTTGTACGGCCGATGACCTTCAAACATATCGCAGATTACACC
TTCATCCAGCGCGCGGGGCTTCGGCAGGAAGCTGTGGGT AAGGCAGATTGTTTTCTGC
TTCCAGT GCCAGAAAATGGCGCTTCTGCTCCGGGCTAAGCACTGGGCTGGTACAATTTG
CTGGCAACGTTGTTGCAGTGCATTTTCATGAGAAGTGGGCATCTTCTTTCTTTATGC
CGAAGGTGATGCGCCATTGTAAGAAGTTT CGT GATGTTCACTTTGATCCTGATGCGTTTG
CCACCACTGACGCATTCACTTGAAGTGAATATTTGAACCAGATCGCATTACAGTGATG
CAAACCTGTAAGTAGATTTCTTAATTGTGATGTGATCGAAGTGTGTTGCGG
>NODE 2 length 309 cov 4.686084
ACTGGTCAGTGCGGGTATCCTTGACAATGGCCGATTGGACGTCTGGCGGAT AAGTTTGG
TCGACTGCTGGTGTGCGTGTTCAGGCTTTTGTGTCATTCTCGGCAGTATCGCGATGCT
TAGCCAGGCGGCGATGGCCCCAGCGTTATTCATCCTCGGTGCCGCTGGCTTTACGCTATA
TCCGGTGGCGATGGCATGGGCTTGCAGAGAAAGTTGAACATCATCAACTGGTGGCGATGAA
CCAGGCCCTTACTGTTGAGCTATACTGTGGGAAGTCTGCTTGGCCCGTCATTTACCGCTAT
GCTAATGCAGAATTTCTCCGATAATTTATTGTT
>NODE 3 length 101 cov 3.346535
AGCGCATGAGCGCGCAGCGCCGCGTTACGTGGTGCATCAGCATGATGTTGGCCGGAGAG
TACAGAGACTCCCCTTCATCCATGATGCCCTCTTTCACCAGCAGTTCTTCAATCATCACC
```

FastqSize \approx GenomeSize x Coverage x 2

At least 300 MB per genome

FastaSize for *E. coli* contigs

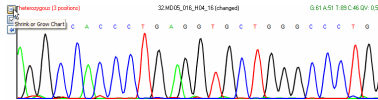
~10 MB

Need for long contigs to investigate the presence of interesting genetic features by blast analysis

7-genes MLST

Conventional Sanger sequencing

PCR, sequencing, electropherograms analysis



Uploading sequences on a webserver to obtain the corresponding alleles and STs

THE UNIVERSITY OF
WARWICK

NGS-derived

Direct upload of WGS contigs on a webserver
(e.g. ARIES or CGE)

Alleles are directly retrieved through blastn comparison with pre-installed database of alleles from University of Warwick with pre-compiled pipelines

Escherichia coli

Ubiquitous
Commensal
Some cause infections

- ▶ Urinary tract infections
- ▶ Sepsis/meningitis
- ▶ Diarrhoea-associated infections

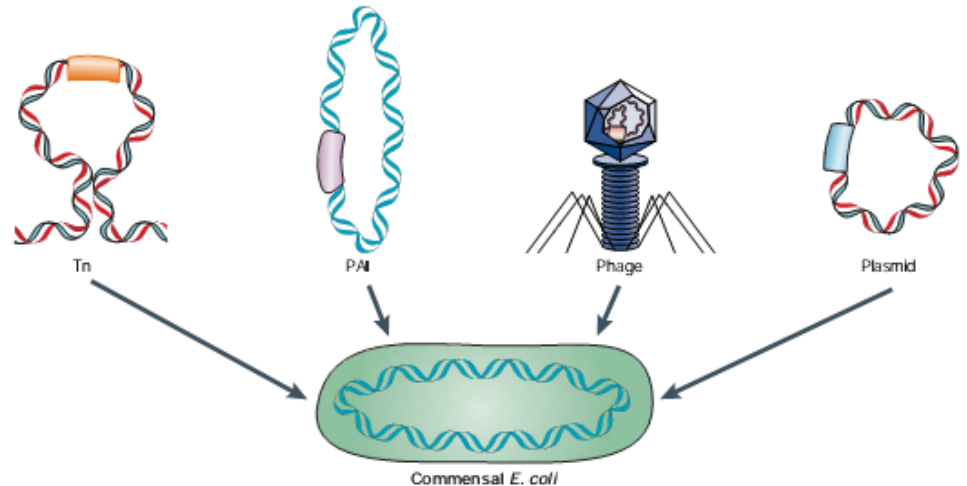


EPEC
ETEC
EIEC
EAEC
DAEC
STEC

Enteropathogenic *E. coli*
Enterotoxigenic *E. coli*
Enteroinvasive *E. coli*
Enteroblastogenic *E. coli*
Enterohemorrhagic *E. coli*
Enterocytotoxic *E. coli*
Shiga-toxin producing *E. coli*

Virulence determinants

- Toxins
- Factors involved in colonization

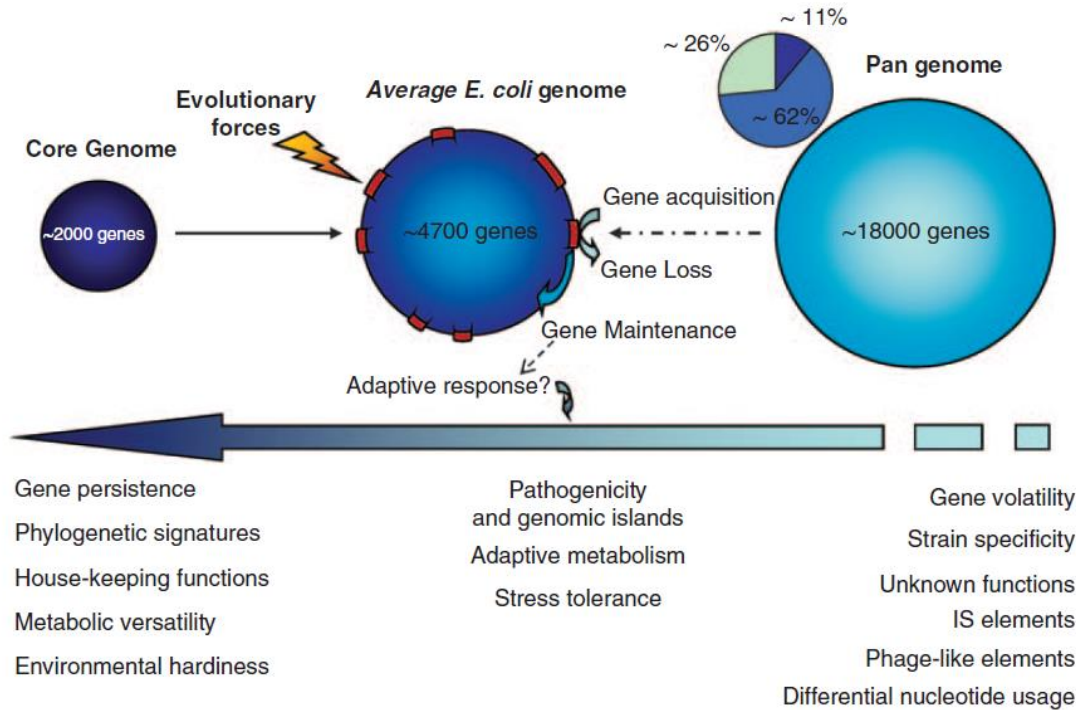


Kaper, Nataro *Nature Reviews* 2004

The *E. coli* pangenome

Genomic plasticity

Huge pangenome



Van Elsas J.D. et al., 2011

Pangenome

Whole genome

Core genome

Accessory genome

Housekeeping

genome



ARIES: A Galaxy-based workspace for intensive data analyses

Galaxy / ARIES Using 244.1 G

Tools

search tools

--- COMMON TOOLS ---

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Statistics
- Graph/Display Data
- GraphAn

---HREVPAP TOOLS---

HReVAP

---NGS TOOLS---

- In Silico PCR
- E coli typing
- NGS: Assembly
- NCBI Blast
- Manipulation
- Gene Annotation
- FASTA/FASTQ manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: BED Tools
- NGS: QC and manipulation
- Operate on Genomic Intervals

---METAGENOMICS TOOLS---

- MetaGenomics
- MetaPhlan2
- Commet

Istituto Superiore di Sanita'

ARIES - Advanced Research Infrastructure for Experimentation in Genomics - Galaxy Instance at ISS

Tweet di @ARIES_GENOMICS

Please read our [disclaimer](#) before using ARIES.

Warning: - FTP is now available for data upload at ariesftp.iss.it (explicit FTP over TLS)

- Take an interactive tour: [Galaxy UI](#) [History](#) [Scratchbook](#)

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by [The Galaxy Team](#) with the support of many contributors. The Galaxy Project is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Johns Hopkins University](#).

History

search datasets

O26comparison
1179 shown, 2160 deleted, 378 hidden
35 GB

1st of 3 pages

- 3565: SNPs_all_matrix.fasta
- 3564: tree_tipAlleleCounts.ML.tre
- 3563: tree_AlleleCounts.ML.NodeLabel.tre
- 3562: tree_AlleleCounts.ML.tre
- 3561: tree.ML.tre
- 3560: tree_tipAlleleCounts.parsimony.tre
- 3559: tree_AlleleCounts.parsimony.No deLabel.tre
- 3558: tree_AlleleCounts.parsimony.tre
- 3557: tree.parsimony.tre
- 3556: O26_paper_Acilia
- 3555: aqMLST Log File
- 3554: aqMLST New Alleles File
- 3553: aqMLST



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



ARIES geographical spread

**Total 76
users**

**Including National
Reference
Laboratories from 13
Member States in EU**

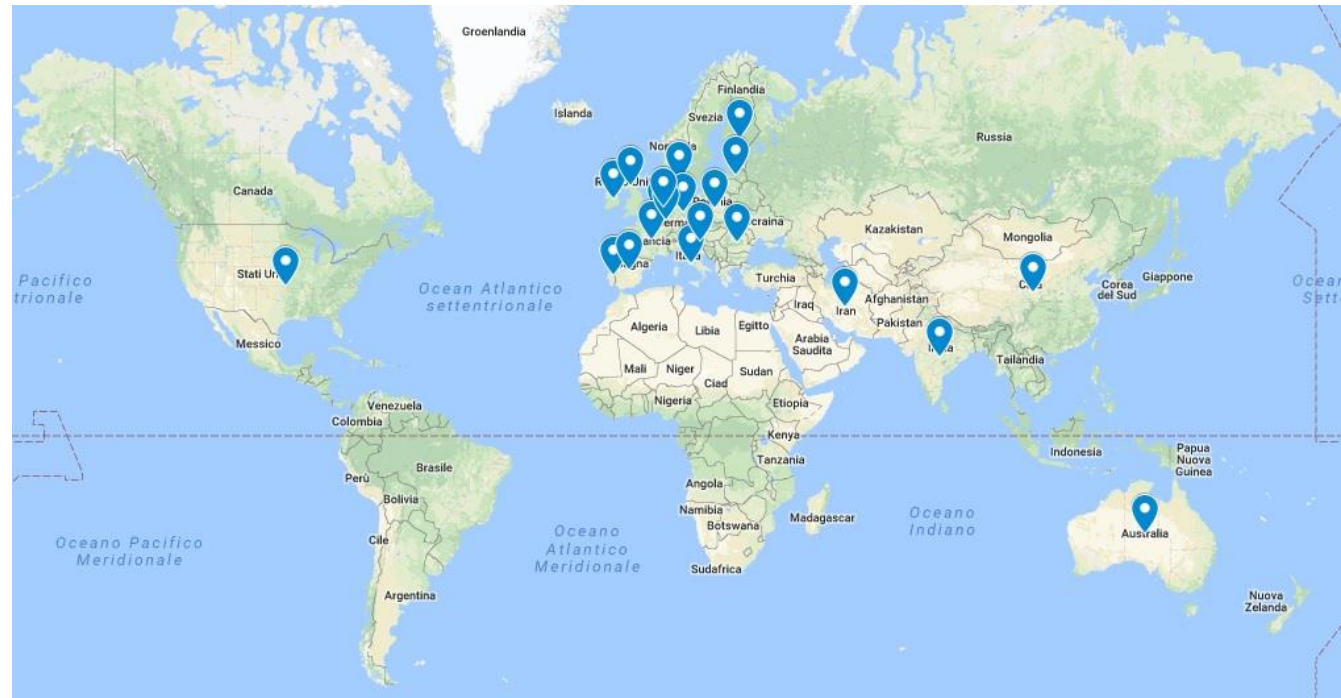
Belgium
Denmark
Finland
Germany
Ireland
Italy

Latvia
The Netherlands
Poland (2 NRLs)
Portugal
Slovenia
United Kingdom

Other users from
Europe:
France, Luxemburgh,
Romania, Spain

Users from
outside Europe:

Australia, China,
India, Iran, USA



NGS data analysis vision: user - friendly user interface

```
a5_pipeline.pl
Getopt::Long::Configure(qw{no_auto_abbrev no_ignore_case_always_pass_through});
my $start = 1;
my $end = 5;
my $sproc = 0;
my $sdebug = 0;
my $metagenome = 0;
my $threads = 4;
my $adapter = dirname(abs_path($0))."/../adapter.fasta";
GetOptions( 'begin=i' => $start,
            'end=i' => $end,
            'preprocessed' => $sproc,
            'debug' => $sdebug,
            'metagenome' => $metagenome,
            'threads=i' => $threads,
            'adapter=s' => $adapter);

die $usage if (@ARGV < 2);

$AVAILMEM = get_availmem();

# Check that java is available and in the path
#
my $java = `which java`;
die "Unable to find a java runtime environment. The A5 pipeline requires java 6 or later. Please ensure that java is installed and in the \SPATH" unless
length($java)>1;
# OS compatibility checks
if($? !~ /darwin/ && $? !~ /linux/){
    die "Sorry, A5-misra only works on Linux and OS X platforms\n";
}
if($? ~ /darwin/){
    my $sga_test = system("SDIR/sga.2> /dev/null > /dev/null");
    if($sga_test){
        die "Sorry, A5-misra does not work on your system.\nThe SGA component of A5-misra requires OS X 10.6 or later\n";
    }
}

# check for spaces in file and path names -- these are not supported
my $message = "contains unsupported characters such as quotes, spaces, or &:%?*<<$";

my $libfile = $ARGV[0];
my $OUTBASE = $ARGV[1];

# check whether command-line input was FastQ files or a library file
if(@ARGV==3){
```

A5 pipeline A5 is a pipeline for assembling DNA sequence data generated on the Illumina sequencing platform. (Galaxy Version 20150522) Options

First read file in fastq format

Second read file in fastq format

A5 is a pipeline for assembling DNA sequence data generated on the Illumina sequencing platform. There are many situations where A5 is not the right tool Illumina data with certain characteristics. A5 will likely not work well with Illumina reads shorter than around 80nt, or reads where the base qualities are low homozygous haploid genomes. Use a different assembler for metagenomes and heterozygous diploid or polyploid organisms. Use a different assembler if a warned!



Esempi

- Caratterizzazione di un clone di *Escherichia coli* enteroinvasivi (EIEC) circolante in Europa
- Caratterizzazione del ceppo STEC-EAEC O104:H4 responsabile dell'epidemia in Germania nel 2011
- Utilizzo dell'NGS in metagenomica

Analisi di un patogeno ri-emergente

Aumento di casi di infezioni da *Escherichia coli* enteroinvasivi in Europa

- Epidemia in Italia nel 2012
- Caso isolato in Spagna nel 2012
- Epidemia in UK nel 2014

Sequenza genomica di due isolati mediante Ion Torrent presso l'ISS



Sequenza genomica di un isolato mediante Illumina presso il PHE



Analisi del contenuto genico

Ricerca di determinanti genetici di interesse attraverso allineamenti blast

- Geni di virulenza
- Geni associati al sierotipo
- Geni parte degli schemi MLST
- Presenza di isole di patogenicità

Esistono database di riferimento per i geni di interesse?

Se non esistono: compilazione di databases ad hoc come liste di sequenze geniche in formato fasta

Bibliografia EIEC O96:H19

Characterization of an emergent clone of enteroinvasive *Escherichia coli* circulating in Europe.

Michelacci V, Prosseda G, Maugliani A, Tozzoli R, Sanchez S, Herrera-León S, Dallman T, Jenkins C, Caprioli A, Morabito S.

Clin Microbiol Infect. 2016 Mar;22(3):287.e11-9. doi: 10.1016/j.cmi.2015.10.025.

PMID: 26551840

Abstract

Enteroinvasive *Escherichia coli* (EIEC) cause intestinal illness indistinguishable from that caused by *Shigella*, mainly in developing countries. Recently an upsurge of cases of EIEC infections has been observed in Europe, with two large outbreaks occurring in Italy and in the United Kingdom. We have characterized phenotypically and genotypically the strains responsible for these epidemics together with an additional isolate from a sporadic case isolated in Spain. The three isolates belonged to the same rare serotype O96:H19 and were of sequence type ST-99, never reported before in EIEC or *Shigella*. The EIEC strains investigated possessed all the virulence genes harboured on the large plasmid conferring the invasive phenotype to EIEC and *Shigella* while showing only some of the known chromosomal virulence genes and none of the described pathoadaptative mutations. At the same time, they displayed motility abilities and biochemical requirements resembling more closely those of the non-pathogenic *E. coli* rather than the EIEC and *Shigella* strains used as reference. Our observations suggested that the O96:H19 strains belong to an emerging EIEC clone, which could be the result of a recent event of acquisition of the invasion plasmid by commensal *E. coli*.

Clinical Microbiology and Infection © 2015 European Society of Clinical Microbiology and Infectious Diseases. Published by Elsevier Ltd. All rights reserved.



Escherichia coli produttori di Shiga-tossina

Patogeni zoonotici

Trasmissione attraverso ingestione di alimenti contaminati



Manifestazioni cliniche intestinali: Assenza di sintomi – Diarrea - Colite emorragica

Manifestazioni cliniche sistemiche: Sindrome Emolitico Uremica (10%): causa principale di danno renale nell'infanzia. Richiede dialisi

Più di 180 sierogruppi circolanti. Solo alcuni sono associati frequentemente con casi di malattia grave (**TOP 5:** O157, O26, O111, O103, O145)

Grande variabilità dovuta alla grande plasticità genomica: geni **stx** veicolati da batteriofagi e **fattori di colonizzazione** fanno parte del genoma accessorio

La Shiga-tossina è anche detta Verocitotossina, quindi i geni *stx* sono anche detti *vtx* e gli STEC sono anche detti VTEC

Maggio 2011: epidemia da STEC-EAEC O104:H4

ALLARME

Pauro in Germania: «germe killer» fa le prime vittime

Quattro morti e oltre 600 casi per un'epidemia causata dal batterio fecale, Escherichia Coli

CORRIERE DELLA SERA



GERMANIA

L'E.coli adesso spaventa l'Ue Oms: "Una variante mai vista prima"

Identificato il ceppo: è nuovo, altamente tossico e resistente ad alcuni antibiotici. In Germania si parla di oltre 1500 casi, più i quasi 500 in altri Paesi Ue, ma è molto difficile fare stime precise. La Russia blocca le importazioni di verdura dall'Europa e Bruxelles protesta. Scagionati i cetrioli, la Spagna chiede i danni all'Unione europea

la Repubblica.it

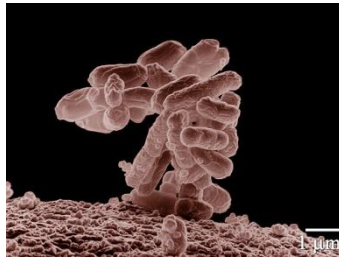


4033 casi

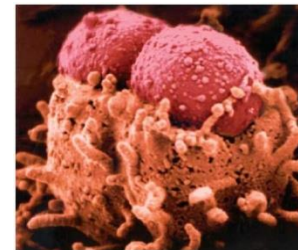
901 SEU

50 decessi

- Sierotipo **O104:H4** (raro)
- Adesione a pila di mattoni (typical of **EAEC**)



Enteroaggregative *E. coli*



Attaching/Effacing *E. coli*

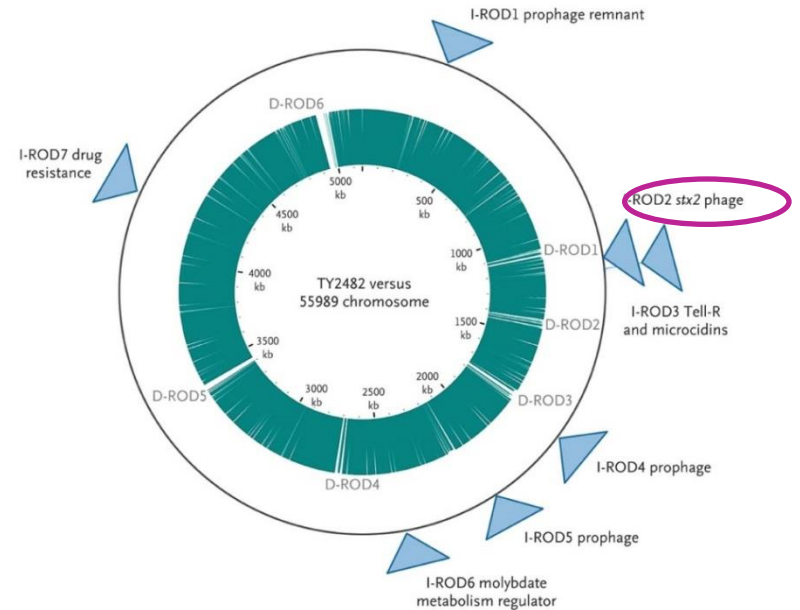
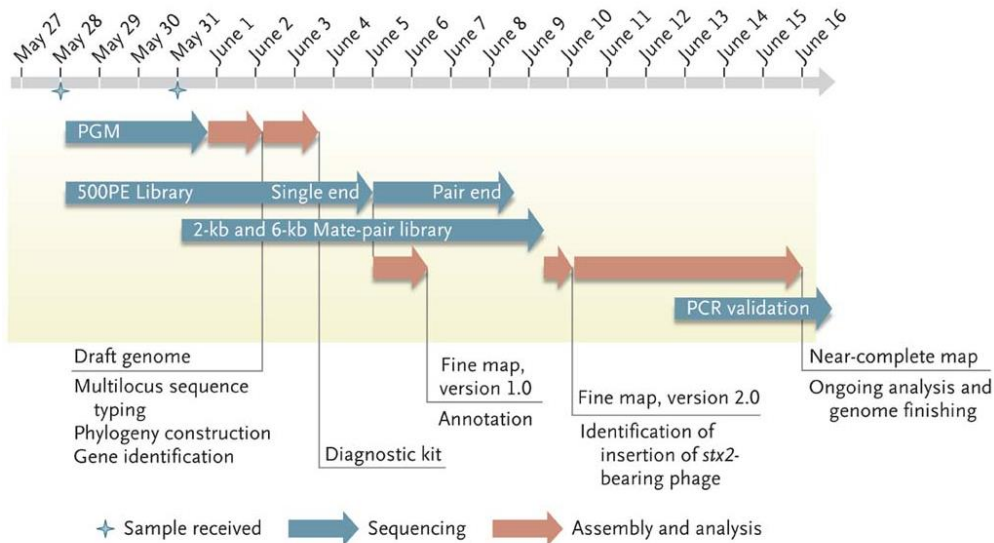


Istituto Superiore di Sanità, Dip. Sanità Pubblica Veterinaria e Sicurezza Alimentare,
Laboratorio Europeo e Nazionale di Riferimento per *E. coli*



Sequenziamento del genoma del ceppo O104:H4

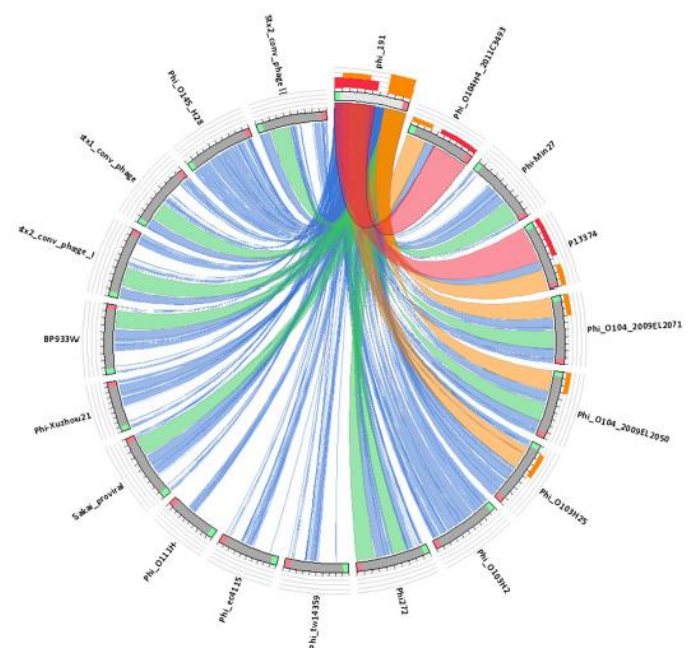
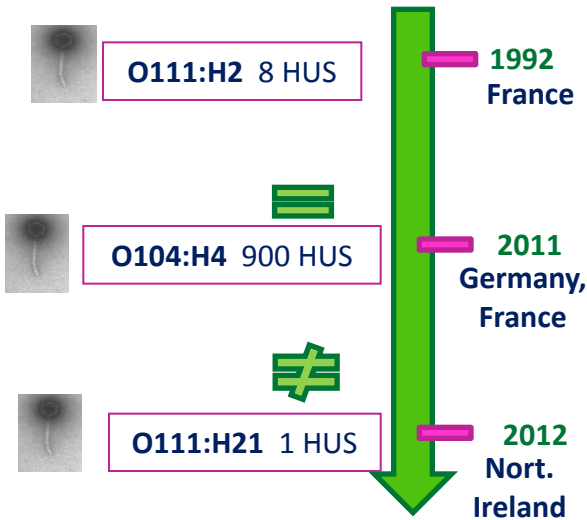
Rohde H et al. N Engl J Med 2011;365:718-724



- Rapido sviluppo di metodi molecolari per identificarne la presenza (geni target: *stx*, *WZX*_{O104}, *fliC*_{H4})
- Studio delle dinamiche evolutive che ne hanno permesso l'emergenza

Genetic Element	Notable Features or Functions	Size or 55989 Coordinates*
Plasmid		
pESBL TY2482	Incl1 plasmid, homologous to pEC_Bactec carrying <i>bla</i> CTX-M-15	88 kb
pAA TY2482	Plasmid encoding aggregative adherence fimbriae I	76 kb
pG2011 TY2482	Plasmid with no obvious phenotype	1.5 kb

Fagi veicolanti geni *stx* in ceppi EAEC

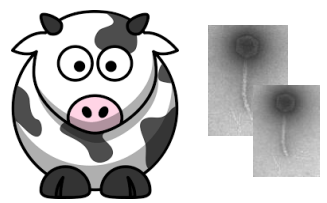


Stesso fago *stx2* conservato in due ceppi isolati a 20 anni di distanza

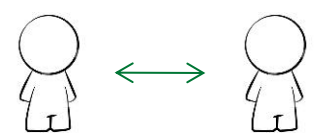
Almeno DUE fagi *stx2* sono in grado di infettare stabilmente ceppi EAEC

Un frammento genico codificante le fibre caudali è conservato in tutti i fagi *stx2* da EAEC

Ipotesi sull'emergenza di ceppo STEC-EAEC



Sorgente degli STEC e dei fagi *stx*



Sorgente degli EAEC



Sovrapposizione delle due sorgenti



Bibliografia STEC-EAEC O104:H4

Whole genome sequence comparison of vtx2-converting phages from Enteroaggregative Haemorrhagic *Escherichia coli* strains.

Grande L, Michelacci V, Tozzoli R, Ranieri P, Maugliani A, Caprioli A, Morabito S.
BMC Genomics. 2014 Jul 8;15:574. doi: 10.1186/1471-2164-15-574
PMID: 25001858

Abstract

Background: Enteroaggregative Haemorrhagic *E. coli* (EAHEC) is a new pathogenic group of *E. coli* characterized by the presence of a *vtx2*-phage integrated in the genomic backbone of Enteroaggregative *E. coli* (EAggEC). So far, four distinct EAHEC serotypes have been described that caused, beside the large outbreak of infection occurred in Germany in 2011, a small outbreak and six sporadic cases of HUS in the time span 1992–2012. In the present work we determined the whole genome sequence of the *vtx2*-phage, termed Phi-191, present in the first described EAHEC O111:H2 isolated in France in 1992 and compared it with those of the *vtx*-phages whose sequences were available.

Results: The whole genome sequence of the Phi-191 phage was identical to that of the *vtx2*-phage P13374 present in the EAHEC O104:H4 strain isolated during the German outbreak 20 years later. Moreover, it was also almost identical to those of the other *vtx2*-phages of EAHEC O104:H4 strains described so far. Conversely, the Phi-191 phage appeared to be different from the *vtx2*-phage carried by the EAHEC O111:H21 isolated in the Northern Ireland in 2012. The comparison of the *vtx2*-phages sequences from EAHEC strains with those from the *vtx*-phages of typical Verocytotoxin-producing *E. coli* strains showed the presence of a 900 bp sequence uniquely associated with EAHEC phages and encoding a tail fiber.

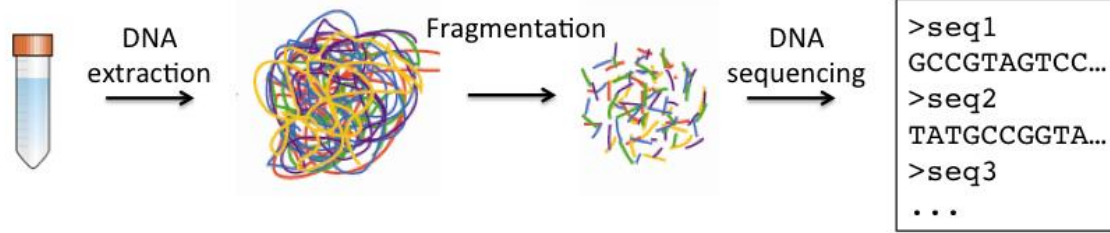
Conclusions: At least two different *vtx2*-phages, both characterized by the presence of a peculiar tail fiber-coding gene, intervened in the emergence of EAHEC. The finding of an identical *vtx2*-phage in two EAggEC strains isolated after 20 years in spite of the high variability described for *vtx*-phages is unexpected and suggests that such *vtx2*-phages are kept under a strong selective pressure.

The observation that different EAHEC infections have been traced back to countries where EAggEC infections are endemic and the treatment of human sewage is often ineffective suggests that such countries may represent the cradle for the emergence of the EAHEC pathotype. In these regions, EAggEC of human origin can extensively contaminate the environment where they can meet free *vtx*-phages likely spread by ruminants excreta.

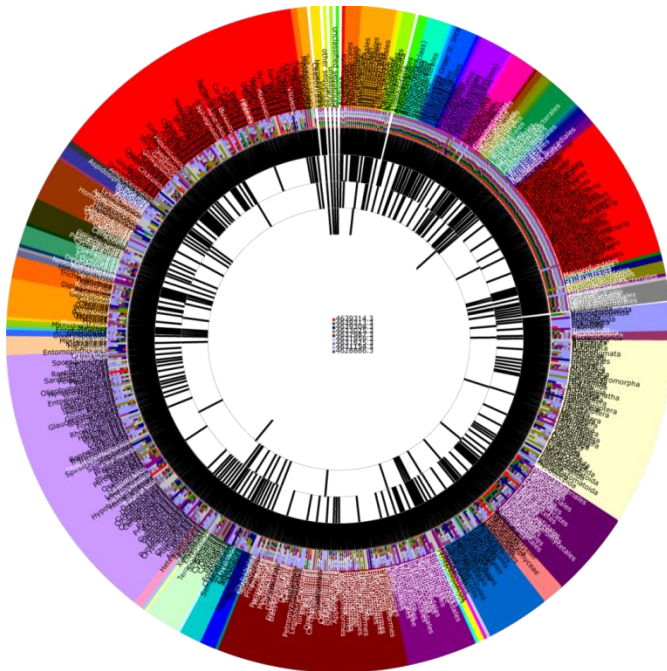
Keywords: Enteroaggregative haemorrhagic *E. coli*, *vtx*-phages, Whole genome sequence, Tail fibers



Applicazioni dell'NGS in Metagenomica



Abbondanza tassonomica relativa (MG-RAST)



Resistenza agli antimicrobici (Diamond)



Bibliografia Metagenomica/1

Comparative analysis of metagenomes of Italian top soil improvers.

Gigliucci F, Brambilla G, Tozzoli R, **Michelacci V**, Morabito S.

Environ Res. 2017 May;155:108-115. doi: 10.1016/j.envres.2017.02.004.

PMID: 28214713

A B S T R A C T

Biosolids originating from Municipal Waste Water Treatment Plants are proposed as top soil improvers (TSI) for their beneficial input of organic carbon on agriculture lands. Their use to amend soil is controversial, as it may lead to the presence of emerging hazards of anthropogenic or animal origin in the environment devoted to food production. In this study, we used a shotgun metagenomics sequencing as a tool to perform a characterization of the hazards related with the TSIs. The samples showed the presence of many virulence genes associated to different diarrheagenic *E. coli* pathotypes as well as of different antimicrobial resistance-associated genes. The genes conferring resistance to Fluoroquinolones was the most relevant class of antimicrobial resistance genes observed in all the samples tested. To a lesser extent traits associated with the resistance to Methicillin in *Staphylococci* and genes conferring resistance to Streptothricin, Fosfomycin and Vancomycin were also identified. The most represented metal resistance genes were cobalt-zinc-cadmium related, accounting for 15–50% of the sequence reads in the different metagenomes out of the total number of those mapping on the class of resistance to compounds determinants. Moreover the taxonomic analysis performed by comparing compost-based samples and biosolids derived from municipal sewage-sludges treatments divided the samples into separate populations, based on the microbiota composition. The results confirm that the metagenomics is efficient to detect genomic traits associated with pathogens and antimicrobial resistance in complex matrices and this approach can be efficiently used for the traceability of TSI samples using the microorganisms' profiles as indicators of their origin.

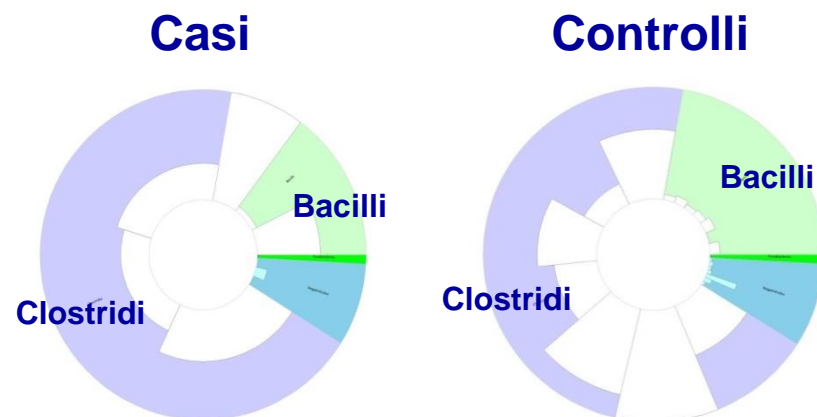


Studio del microbiota in un focolaio epidemico da STEC O26

Confronto con database di sequenze di geni di virulenza: conferma della presenza di STEC O26 *eae stx2a* nei campioni prelevati dai casi

Studio della composizione tassonomica del microbiota intestinale nei casi VS i controlli

A. 9	A. 4	A. 30	A. 40	A. 8	A. 16	A. 14	481-5	A. 32	A. 41
cif	epeA	senB	ehxA	cif	cba	cif		katP	mhcF
eae	espP		espP	eae	cma	eae		lpfA	tsh
efa1	gad		katP	efa1	gad	efa1			
ehxA	iha		mhcC	ehxA	ireA	ehxA			
espA			pic	espA	iss	espA			
espB			senB	espB	lpfA	espB			
espJ			toxB	espF	pic	espF			
espP				espJ	vat	espJ			
gad				espP		espP			
iha				gad		gad			
iss				iha		iha			
katP				iss		iroN			
lpfA				katp		iss			
nleA				lpfA		katP			
nleB				nleA		lpfA			
nleC				nleB		nleA			
stx2a				nleC		nleB			
tir				stx2a		nleC			
toxB				tir		stx2a			
				toxB		tir			
						toxB			



Identificazione di differenze nella rappresentatività di alcune classi di microrganismi che potrebbero svolgere un ruolo protettivo durante l'infezione

Bibliografia Metagenomica/2

Metagenomic Characterization of the Human Intestinal Microbiota in Fecal Samples from STEC-Infected Patients

Gigliucci F, von Meijenfeldt FAB, Knijn A, **Michelacci V**, Scavia G, Minelli F, Dutilh BE, Ahmad HM, Raangs GC, Friedrich AW, Rossen JWA, Morabito S. *Front Cell Infect Microbiol.* 2018 Feb 6;8:25.
PMID: 29468143

Metagenomic Characterization of the Human Intestinal Microbiota in Fecal Samples from STEC-Infected Patients

Federica Gigliucci^{1,2*}, *F. A. Bastiaan von Meijenfeldt*³, *Arnold Knijn*¹, *Valeria Michelacci*¹, *Gaia Scavia*¹, *Fabio Minelli*¹, *Bas E. Dutilh*^{3,4}, *Hamideh M. Ahmad*⁵, *Gerwin C. Raangs*⁵, *Alex W. Friedrich*⁵, *John W. A. Rossen*⁵ and *Stefano Morabito*¹

¹ Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanità, Viale Regina Elena, Rome, Italy, ² Department of Sciences, University Roma Tre, Rome, Italy, ³ Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, Netherlands, ⁴ Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, Nijmegen, Netherlands, ⁵ Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

The human intestinal microbiota is a homeostatic ecosystem with a remarkable impact on human health and the disruption of this equilibrium leads to an increased susceptibility to infection by numerous pathogens. In this study, we used shotgun metagenomic sequencing and two different bioinformatic approaches, based on mapping of the reads onto databases and on the reconstruction of putative draft genomes, to investigate possible changes in the composition of the intestinal microbiota in samples from patients with Shiga Toxin-producing *E. coli* (STEC) infection compared to healthy and healed controls, collected during an outbreak caused by a STEC O26:H11 infection. Both the bioinformatic procedures used, produced similar result with a good resolution of the taxonomic profiles of the specimens. The stool samples collected from the STEC infected patients showed a lower abundance of the members of *Bifidobacteriales* and *Clostridiales* orders in comparison to controls where those microorganisms predominated. These differences seemed to correlate with the STEC infection although a flexion in the relative abundance of the *Bifidobacterium* genus, part of the *Bifidobacteriales* order, was observed also in samples from Crohn's disease patients, displaying a STEC-unrelated dysbiosis. The metagenomics also allowed to identify in the STEC positive samples, all the virulence traits present in the genomes of the STEC O26 that caused the outbreak as assessed through isolation of the epidemic strain and whole genome sequencing. The results shown represent a first evidence of the changes occurring in the intestinal microbiota of children in the course of STEC infection and indicate that metagenomics may be a promising tool for the culture-independent clinical diagnosis of the infection.



Conclusioni e prospettive

Il sequenziamento del genoma completo permette di ottenere dati utili alla caratterizzazione e tipizzazione dei microrganismi patogeni

- Base per lo **sviluppo di metodi** per l'identificazione di un patogeno
- Possibilità di estrarre informazioni utili per **capire l'origine e l'emergenza di sottotipi patogeni**
- Possibilità di confrontare ceppi patogeni ed identificarne le **relazioni filogenetiche**
 - attraverso studio degli SNPs nel genoma totale
 - per confronto del contenuto genico – necessità di database di geni e alleli
- Possibile applicazione della **metagenomica** nella diagnosi delle infezioni e nello studio di campioni complessi per investigare nuove possibili sorgenti di infezione

Ringraziamenti



Stefano Morabito

Alfredo Caprioli

Rosangela Tozzoli

Fabio Minelli

Antonella Maugliani

Paola Chiani

Gaia Scavia

Clarissa Ferreri

Federica Gigliucci

Silvia Arancia

Arnold Knijn



Galaxy / ARIES - ISS



Aries Group @ARIES_GENOMICS



Grazie per l'attenzione!

