



## **Séquençage de Nouvelle Génération et épidémiologie microbienne hospitalière**

# **L'expérience de l'Hôpital NECKER - enfants malades**

**SP14 - Session SFM : Apport pratique des techniques de séquençage de nouvelle génération  
dans l'investigation des phénomènes épidémiques**

Session parallèle Vendredi 8 juin 2018 11:30 - 12:00



# DÉCLARATION DE LIENS D'INTÉRÊTS

**Conférencier: Anne, JAMET, Paris**

**Je n'ai pas de lien d'intérêt potentiel à déclarer**

# La microbiologie clinique @Necker

- L'hôpital Necker
  - 400 lits de pédiatrie (médecine, chirurgie, réa, urgences)
  - 200 lits pour le pôle Adulte (hémato, néphro, SMIT)
- Le laboratoire
  - Centre de référence Mucoviscidose



# Le NGS en microbiologie @Necker

- Premiers pas avec un 454 en 2013 (Roche)
- Acquisition d'un MiniSeq en 2017 (Illumina)

2013



Anne Jamet @DrAnneJamet · 26 Sep 2017  
MiniSeq is in the lab !

2017








# Le séquenceur de “pailleasse”

- Notre choix:
  - La plateforme la plus utilisée
  - Le séquenceur le moins cher
  - Le petit “débit” -> pas de délai
  - Séquençage d’une souche bactérienne
    - Environ 100 euros
    - De la souche à la séquence brute en 4 jours
    - Analyse (de 1h à plusieurs jours...)

# Choix du séquenceur

Short reads < 300 bp

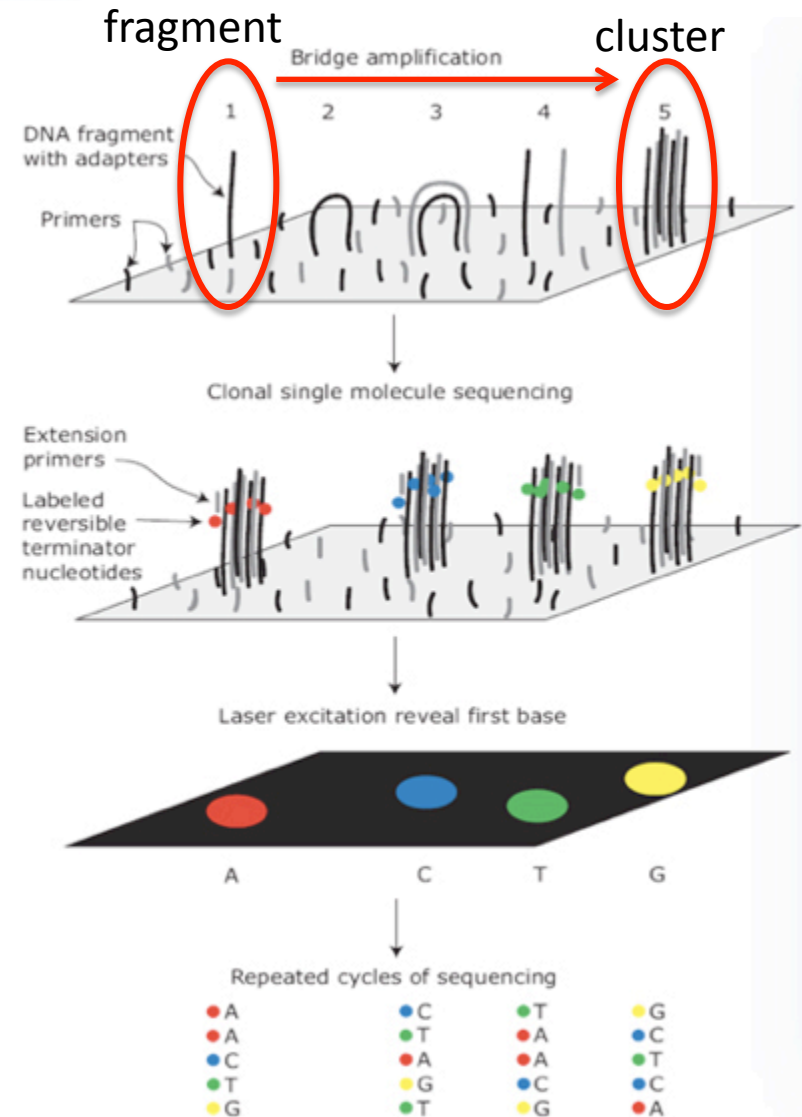
For small labs (benchtop instruments ex: MiSeq) or big genomic centers (HiSeq)

	 MiniSeq System	 MiSeq Series	 NextSeq Series	 HiSeq Series	 HiSeq X Series*
<b>Key Methods</b>	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing.	Small genome, amplicon, and targeted gene panel sequencing.	Everyday exome, transcriptome, and targeted resequencing.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale whole-genome sequencing.
<b>Maximum Output</b>	7.5 Gb	15 Gb	120 Gb	1500 Gb	> 1 To 1800 Gb
<b>Maximum Reads per Run</b>	25 million	25 million†	400 million	5 billion	6 billion
<b>Maximum Read Length</b>	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
<b>Run Time</b>	4–24 hours	4–55 hours	12–30 hours	<1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	<3 days
<b>Benchtop Sequencer</b>	Yes	Yes	Yes	No	No

# Illumina sequencing process



- DNA fragments bind the **flow cell** through the adaptors
- Bridge amplification creates the clusters
- Nucleotides fluorescently-labeled are incorporated
- detected with a CCD camera -> Base calling



# Un peu de terminologie



- **Un « run »**
- **Des « reads »**
  - En général des « short reads » (<300 pb)
  - une séquence d'un fragment d'ADN
  - *paired-end sequencing*: on séquence le fragment aux 2 extrémités mais pas au milieu
- **Des fichiers « FastQ »**
  - Fichier texte: séquence nucléotidique + scores de qualité des nt
  - Q-scores for each base = quality score = prediction of the probability of an error in base calling
  - quality score of Q30 = 1/1,000 predicted to be incorrect

# Complete genome sequencing by NGS

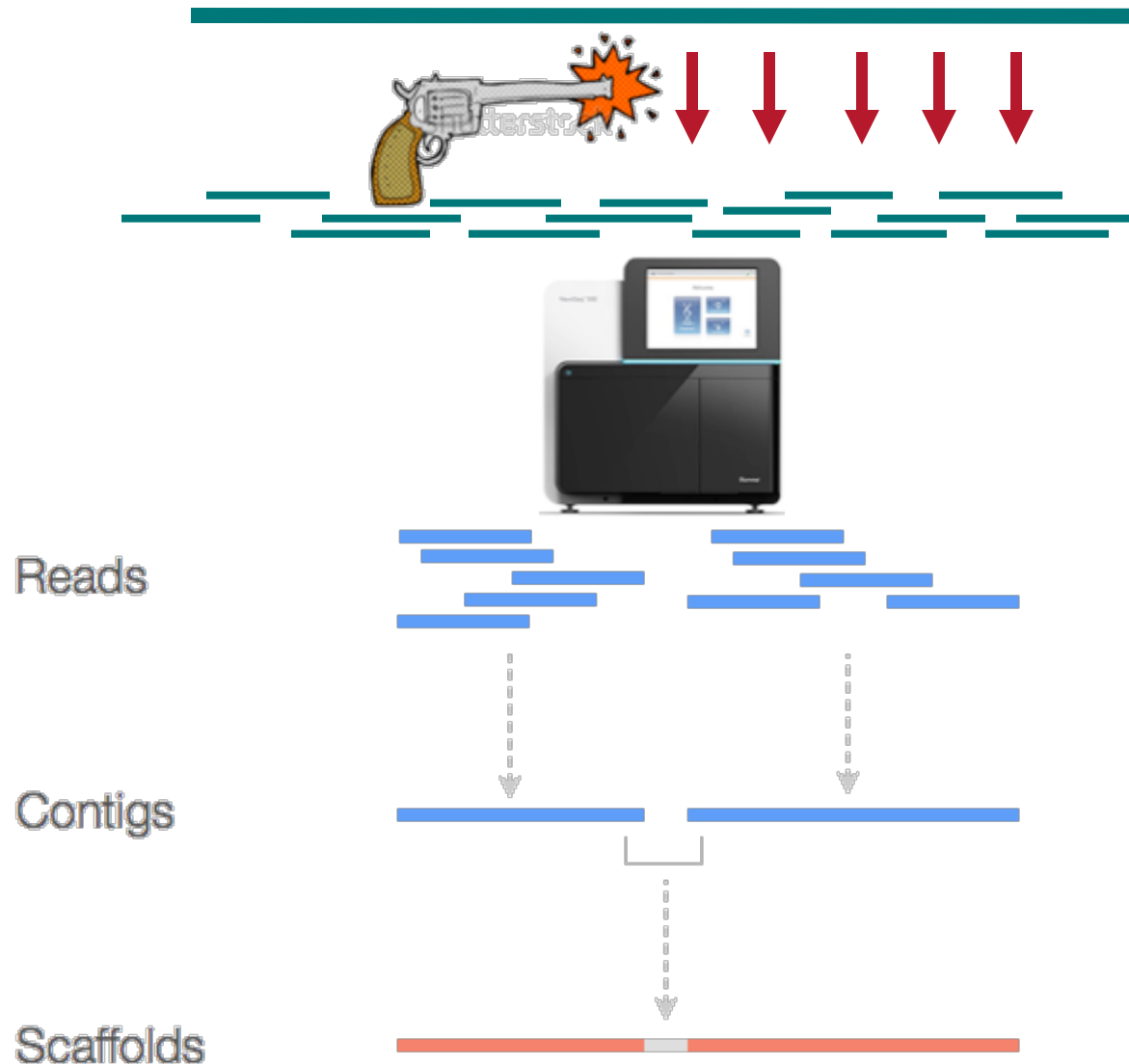
Whole-genome shotgun (WGS) of 1 strain

1- DNA extraction

2- Library preparation

3- Sequencing

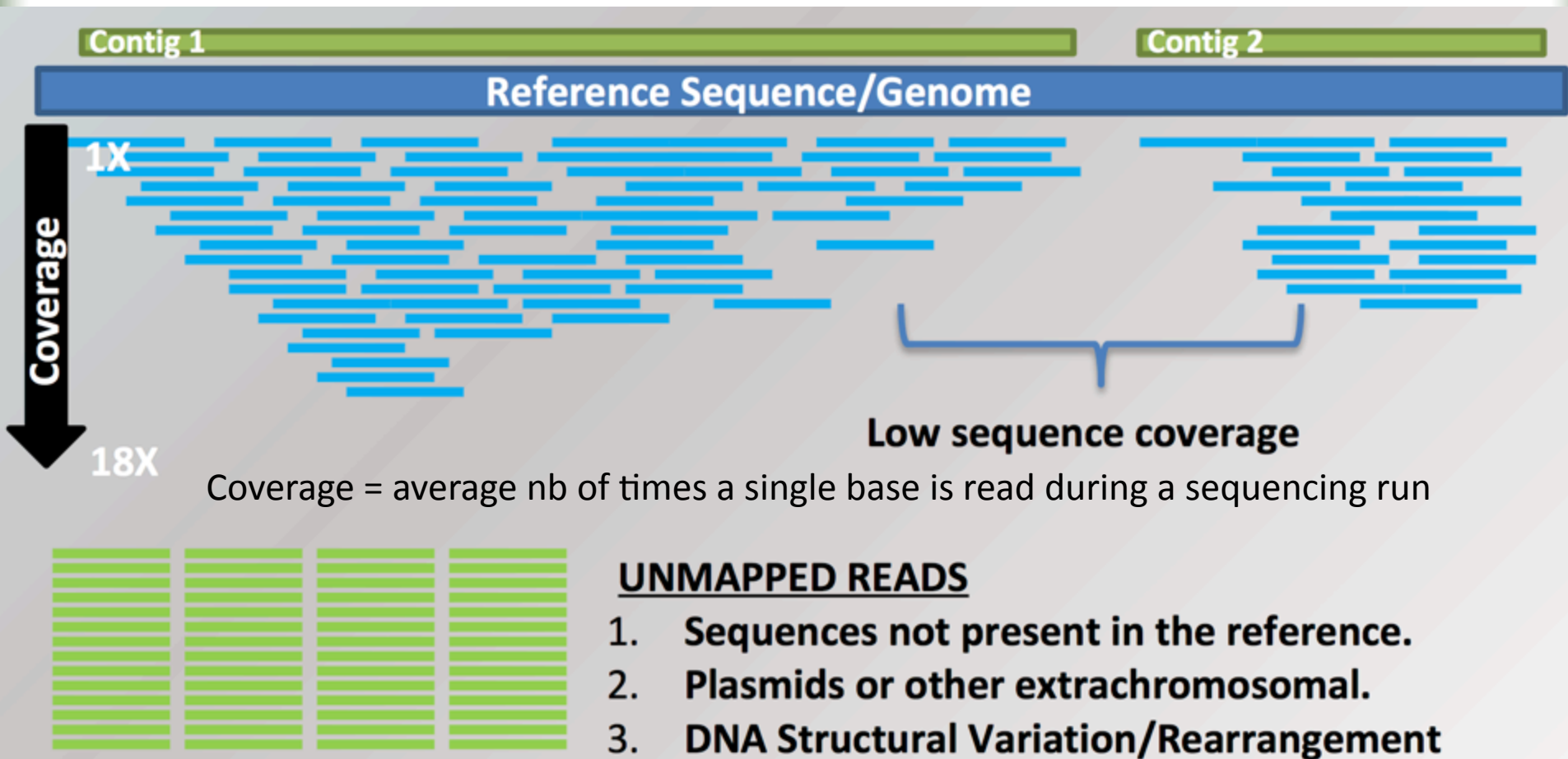
4- Analysis



# Analysis 1

## Mapping of the reads

- **Comparison** to a reference genome or a database of alleles
- Variant calling
  - Identification of the differences (mutations, indels...)



# Analysis 2

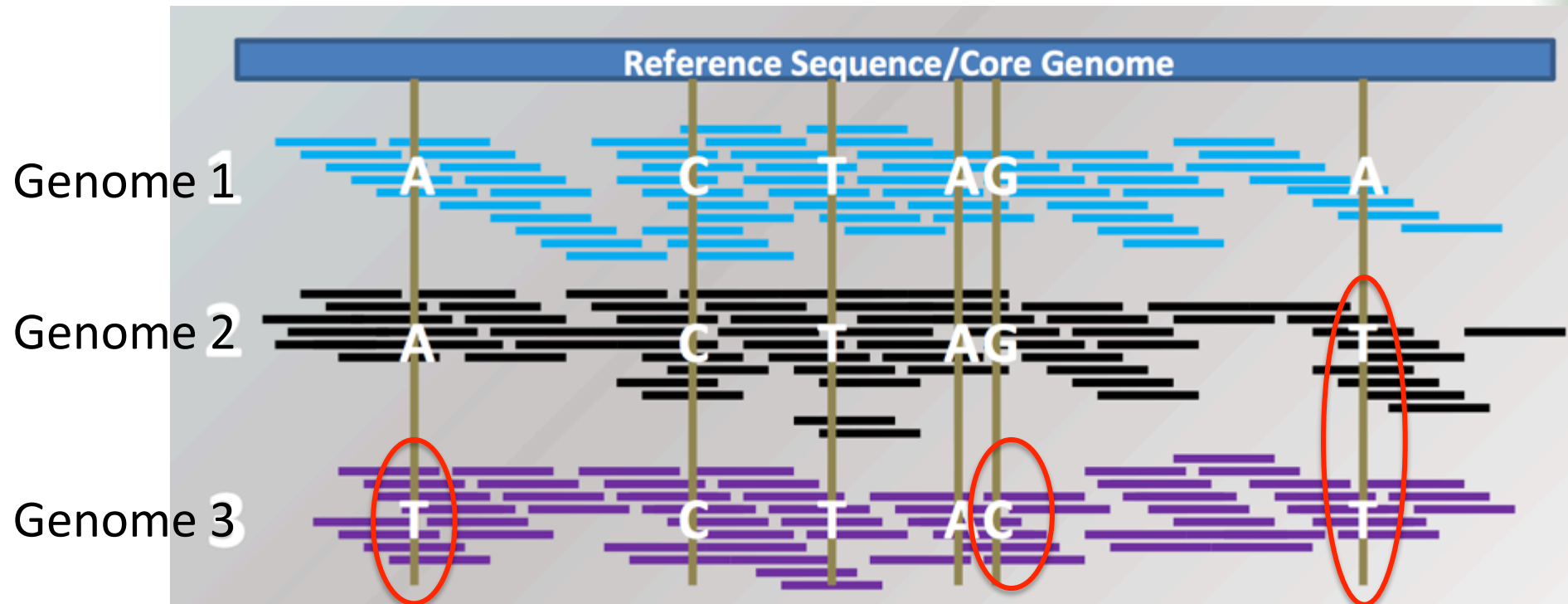
## De novo assembly of the reads++ in microbiology

Without any references



# Comparaison de souches

- 1- Comparer après assemblage des génomes (ex: cgMLST...)
- 2- Comparer sans assembler
  - Identifier les SNPs dans les cores gènes par mapping



# Applications en routine @Necker

- Comparaison de souches si suspicion épidémie
  - *B. cereus*
  - *S. epidermidis*
- Pathogènes spécifiques
  - *M. tuberculosis*
    - Antibiogramme moléculaire de *M. tuberculosis*
    - Typage de *M. tuberculosis*
  - *S. aureus*
    - Liste des gènes de virulence

# Workflow commun pré-analytique

1- Extraire les ADN

2- Préparer les librairies

-> multiplexage d'échantillons

-> ex: *S. aureus* X3 + *M. tuberculosis* X2

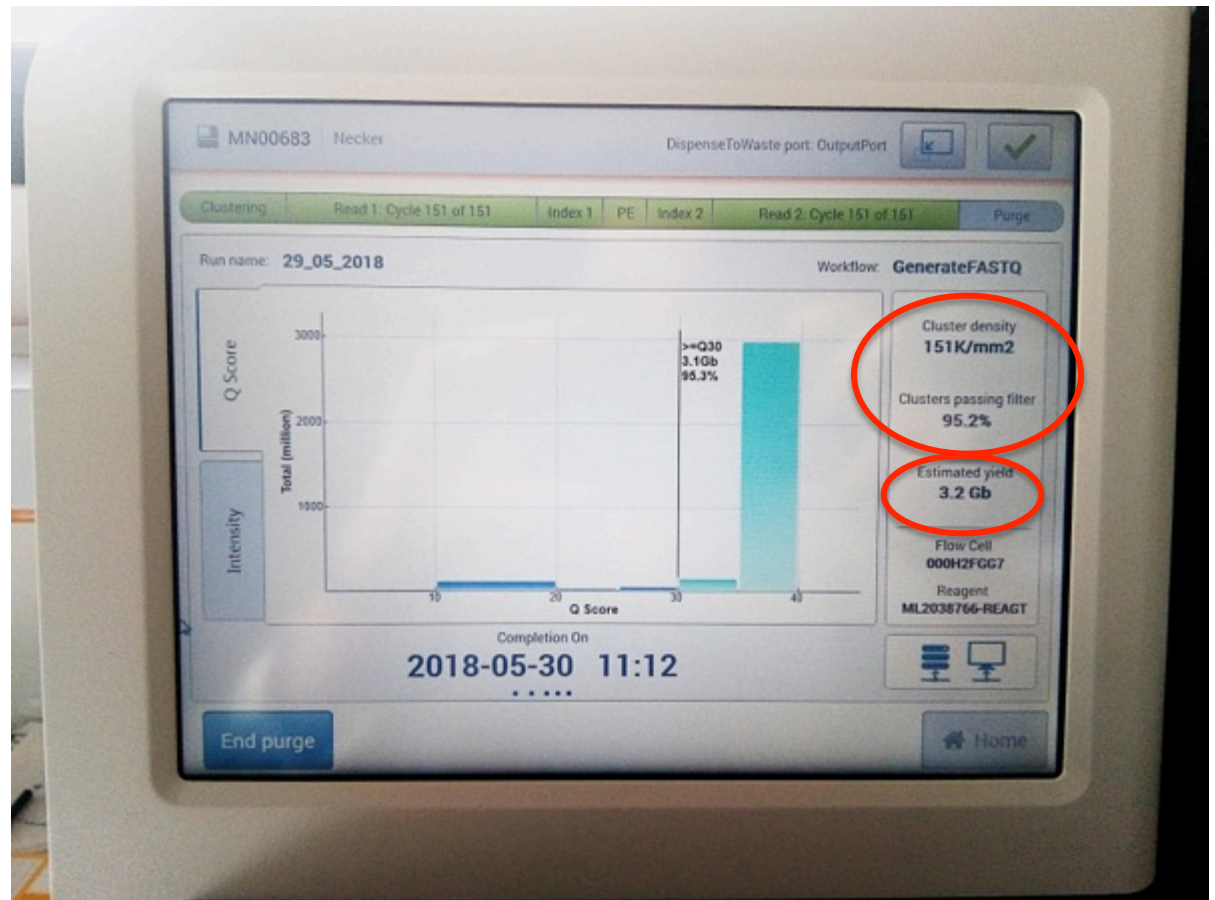
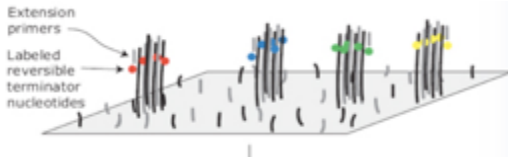
-En pratique 6 à 10 souches selon taille genome

3- Séquençage sur machine

4- Récupération des fichiers bruts “.fastq”

# Evaluer le “run”

- Nombre de “reads” de bonne qualité obtenus
- Bonne repartition entre les souches (multiplexage)



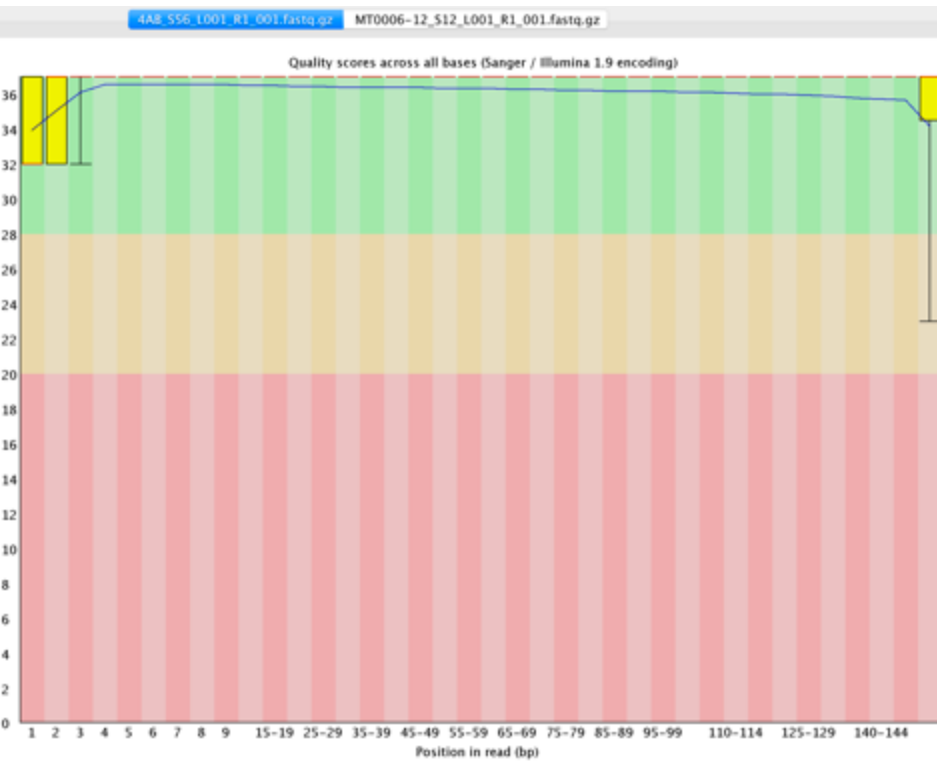
% Q-score  $\geq$  Q30

Q30 is a probability of incorrect base calling of 1 in 1000

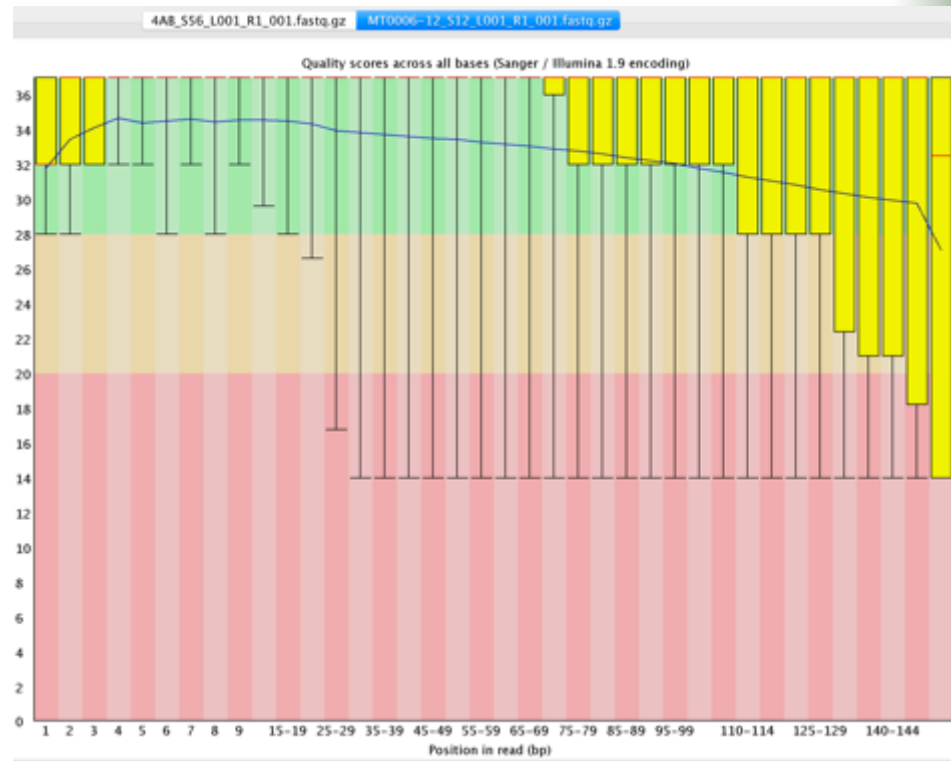
# Evaluer les "reads"

## FastQC

GOOD



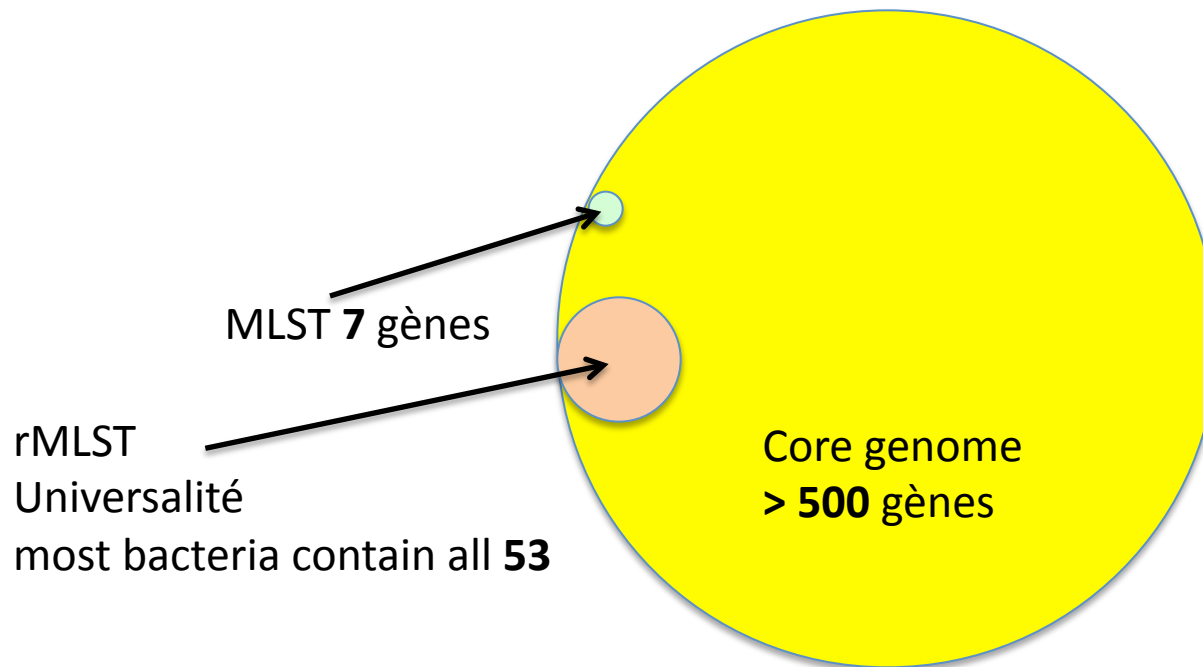
BAD



**Typage moléculaire**



# MLST, ribosomal MLST, core genome MLST



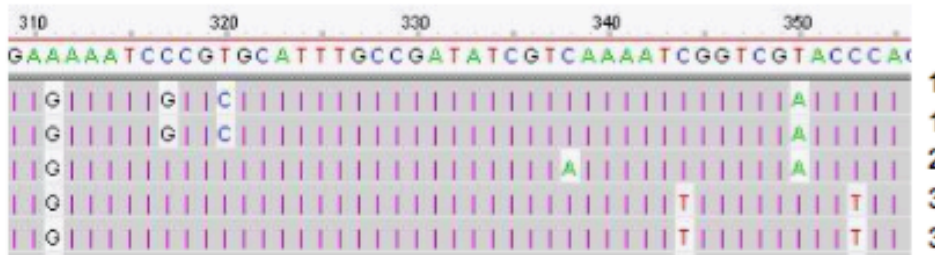
1 recombinaison OU 1 mutation ponctuelle = 1 seul événement = 1 allèle différent

**the number of events, rather than the overall sequence divergence, is a more reliable estimator of relatedness**

# Core genome MLST



**Step 1. Define core genome**  
Core genes



**Step 2. Define variation at core genes**

Alleles of core genes

1	15	12	37	3	16	22	11
1	5	4	24	4	4	1	1
27	28	9	31	1	25	10	15

**Step 3. Define allelic profiles**

Profiles

Avec l'équipe de S. Brisse (IP) **definition cgMLST pour *S. epidermidis* de 699 core-genes**

Slide from S. Brisse

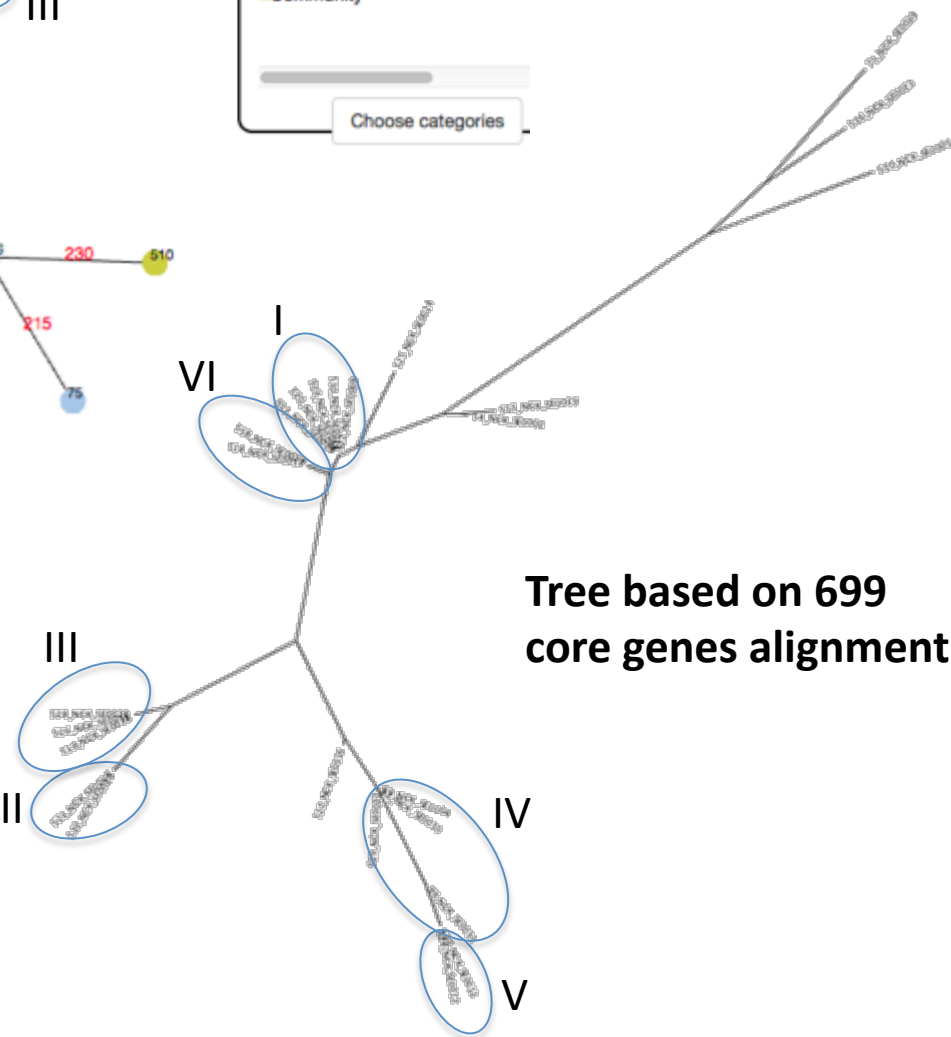
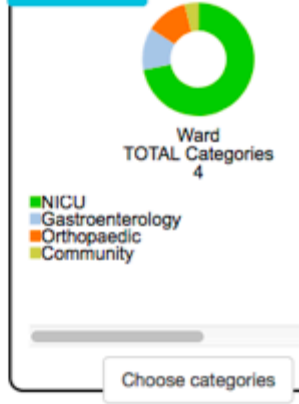
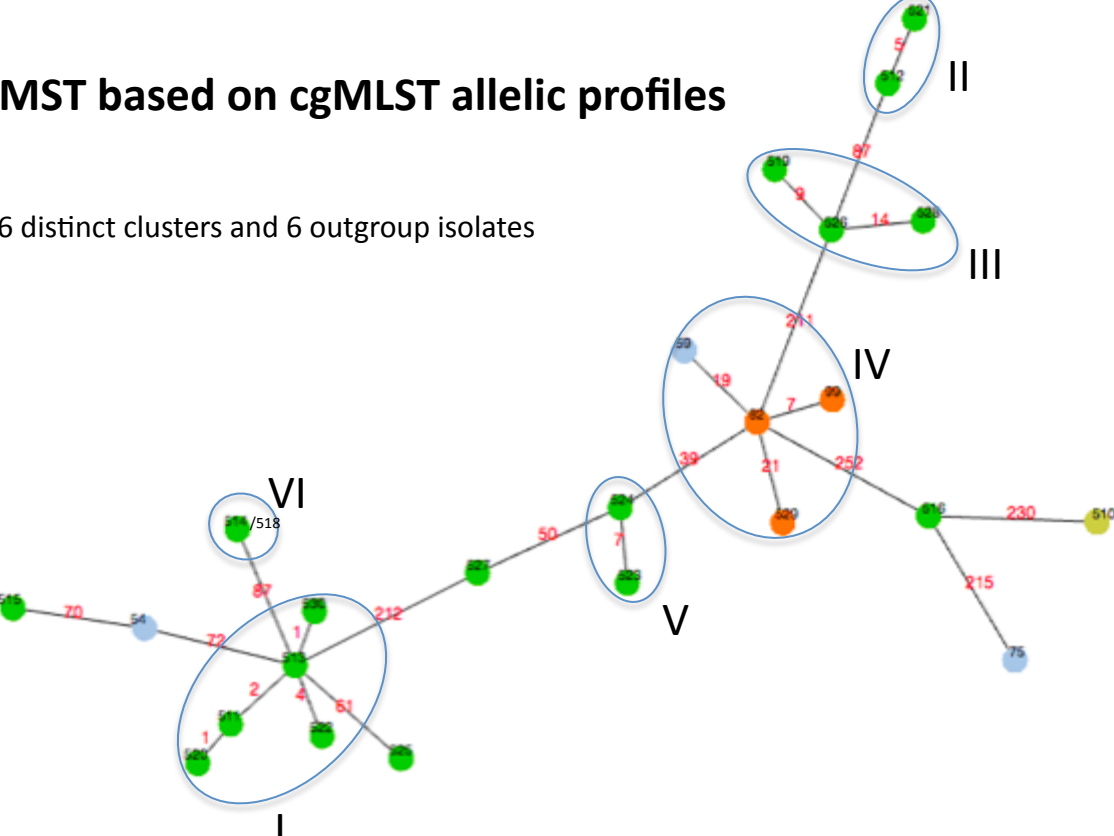
# *S. epidermidis* from Necker

## Test of our cgMLST scheme using 25 isolates from our hospital

- 25 isolates recovered from several wards
    - 18 strains from infants during a **one-year period in a NICU**
    - 6 strains from patients hospitalized in **other wards**
    - 1 strain from a healthy adult
    - 1 isolate /patient except for **3 infants who had 2 isolates** sequenced
- > mainly hospitalized neonates in NICU with MRSE ST2 clone isolated from blood

# MST based on cgMLST allelic profiles

6 distinct clusters and 6 outgroup isolates



Tree based on 699 core genes alignment

- 5 distinct clusters (I,II,III,V,VI) in the NICU across a year
- another cluster = orthopaedic and gastroenterology wards

= good correlation between the cgMLST allelic profile approach and the phylogenetic analysis

# Conclusion cgMLST

- 1<sup>st</sup> cgMLST scheme for *S. epidermidis*
  - integrated it in the publically available BIGSdb server
    - <http://bigsdb.pasteur.fr/epidermidis/>
  - allowing reproducible comparisons among clinical laboratories
- Add virulence and resistance genes

# **Gènes de virulence et typage de *S. aureus***

# Pathogènes spécifiques: *S. aureus*

## 1- Extraire l'ADN de Staphylocoque

- Protocole: lysostaphine

## 2- Utilisation des fichiers bruts “.fastq”

- Typage, virulence et antibiogramme moléculaire
  - Mykrobe predictor
  - CGE
  - **SRST2** (en ligne de commande...)
    - ST, type SCCmec, groupe agr, toxines

## 3- Trimming, assemblage de novo, annotation

- Comparaison de souches

# Pathogènes spécifiques: *S. aureus*

“mapping” des reads bruts (FastQ) contre une database de gènes de virulence et de marqueurs de typage

En ligne de commande

```

Last login: Tue Jun 21 14:35:16 on tty1000
MacBook-Pro:~ Admin$
MacBook-Pro:~/Documents Admin$ ping -c 8 alltime.com
PING alltime.com (46.4.194.4): 56 data bytes
64 bytes from 46.4.194.4: icmp_seq=0 ttl=63 time=0.871 ms
64 bytes from 46.4.194.4: icmp_seq=1 ttl=63 time=0.318 ms
64 bytes from 46.4.194.4: icmp_seq=2 ttl=63 time=0.484 ms
64 bytes from 46.4.194.4: icmp_seq=3 ttl=63 time=0.748 ms
64 bytes from 46.4.194.4: icmp_seq=4 ttl=63 time=0.185 ms
64 bytes from 46.4.194.4: icmp_seq=5 ttl=63 time=0.813 ms
64 bytes from 46.4.194.4: icmp_seq=6 ttl=63 time=0.904 ms
64 bytes from 46.4.194.4: icmp_seq=7 ttl=63 time=0.298 ms

--- alltime.com ping statistics ---
 8 packets transmitted, 8 packets received, 0.0% packet loss
round-trip min/avg/mtt = 0.133/0.51/0.678/0.296/3.141 ms
MacBook-Pro:~/Documents Admin$ ls -la
total 24
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Documents
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Downloads
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Fonts
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Pictures
drwxr-xr-x  4 Admin  staff 320 Mar 22 23:02 Projects
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Services
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Screenshots
drwxr-xr-x  10 Admin  staff  320 Mar 22 23:02 .
drwxr-xr-x  15 Admin  staff 518 Jun 30 13:46 ..
-rw-r--r--  1 Admin  staff 38244 Jun  9 21:06 _05_Store
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Music
drwxr-xr-x  2 Admin  staff  68 Sep 14 2015 Movies
MacBook-Pro:~/Documents Admin$
    
```

## Récapitulatif

Date	2018_04_30					
Sample ID	SA0038					
Sequençage	WGS	Method	Library	Instrument	Cycles	DNA extraction
Run ID	A000H2FMGL	paired-ends 2x150	Nextera XT DNA Library Preparation Kit	MiniSeq MN00683	MID output	DNeasy Blood & Tissue Kit

SRST2 version	v0.2.0
Database	all_genes_chip_Saur_srst2_20170411.fasta
Mykrobe	Predictor S.aureus v0.1.3

Taxonomie	Staphylococcus aureus						
Meti R	absence						
PVL	absence						
ST	arcC	aroE	glpF	gmk	pta	tpi	yqjL
<b>CC121/ST121</b>	6	5	6	2	7	14	5
Groupe agr	IV <i>accessory gene regulator</i>						

## Gènes associés à la résistance aux antibiotiques

Gène	Résultat	Interprétation
<i>mecA</i>	absence	PLP2a confère une résistance à la méthicilline
<i>blaZ</i>	présence	résistance à la pénicilline A (beta-lactamase)
<i>fusB</i>	présence	résistance à l'acide fusidique
<i>mupA</i>	présence	résistance à mupirocin
<i>aadD</i>	présence	résistance aux aminosides

## SCCmec typing (uniquement pour SARM)

SCCmec type	
-------------	--

## Gènes associés à la virulence

ACME locus	absence	<i>arcABCD</i>
------------	---------	----------------

## Enterotoxines

<i>tst1</i>	absence	<i>toxic shock syndrom toxin</i>
<i>sea</i>	absence	<i>enterotoxin A (=entA)</i>
<i>seb</i>	absence	<i>enterotoxin B (=entB)</i>

**Antibiogramme moléculaire  
et typage de *M. tuberculosis***

# Pathogènes spécifiques: *M. tuberculosis*

## 1- Extraire l'ADN de Mycobactéries

- Difficile à lyser ++
- Protocole: à partir colonie ou MGIT -> Fastprep

## 2- Utilisation des fichiers bruts ".fastq"

- Typage et antibiogramme moléculaire
  - en ligne
  - Mykrobe predictor
- Comparaison de souches
  - Mapping contre H37Rv (en ligne de commande...)
  - SNPs: >12 SNPs apart not likely to be related epidemiologically

# Resistance et lineage

Name: 5a99822338467

Sample: 5a99822338467

Drug <sup>1</sup>		Resistance	Supporting Mutations
Isoniazid	<i>katG inhA</i>		
Rifampicin	<i>rpoB</i>	R	rpoB (Gln432Pro)
Ethambutol	<i>EmbCAB</i>		
Pyrazinamide	<i>pncA</i>		

Strain	Lineage
MT0011_S8_L001	lineage4 Euro-American LAM;T;S;X;H None lineage4.1 Euro-American T;X;H None lineage4.1.2 Euro-American (X-type) T;H None lineage4.1.2.1 Euro-American (X-type) T1;H1 RD182
MT0010_S7_L001	lineage4 Euro-American LAM;T;S;X;H None lineage4.3 Euro-American (LAM) mainly-LAM None lineage4.3.1 Euro-American (LAM) LAM9 None
MT0008-2_S4_L001	lineage1 Indo-Oceanic EAI RD239 lineage1.1 Indo-Oceanic EAI3;EAI4;EAI5;EAI6 RD239 lineage1.1.2 Indo-Oceanic EAI3;EAI5 RD239
MT0005-2_S2_L001	lineage4 Euro-American LAM;T;S;X;H None lineage4.1 Euro-American T;X;H None lineage4.1.2 Euro-American (X-type) T;H None lineage4.1.2.1 Euro-American (X-type) T1;H1 RD182
1AB_S12_L001	lineage4 Euro-American LAM;T;S;X;H None lineage4.6 Euro-American T;LAM None lineage4.6.2 Euro-American T;LAM RD726 lineage4.6.2.2 Euro-American (Cameroon) LAM10-CAM RD726

En ligne

TGS-TB et TB profiler

<https://gph.niid.go.jp/tgs-tb/>

<http://tldr.lshtm.ac.uk/>

# Pathogènes spécifiques: *M. tuberculosis*

- Prédiction de la résistance aux ATB
  - *M. tuberculosis* (Rif, INH, Emb)



**MYKROBE  
PREDICTOR  
TB** Se 82.6% Sp 98.5%

En local

**R** RIFAMPICIN

Resistance mutation found: Q432X in gene rpoB  
Resistant allele seen 34 times  
Susceptible allele seen 0 times

**S** SPECIES

*M. tuberculosis* (lineage: European/American)

**SUSCEPTIBLE**

Isoniazid  
Ethambutol  
Quinolones  
Streptomycin  
Amikacin  
Capreomycin  
Kanamycin

**RESISTANT**

Rifampicin

# Remerciements



**Pr X. Nassif**  
**Head of Lab**

**Necker Team**



**Institut Pasteur**

**S. Brisse**

**J. Guglielmini**

