Bacterial sequencing on ONT MinION

Yann Reynaud

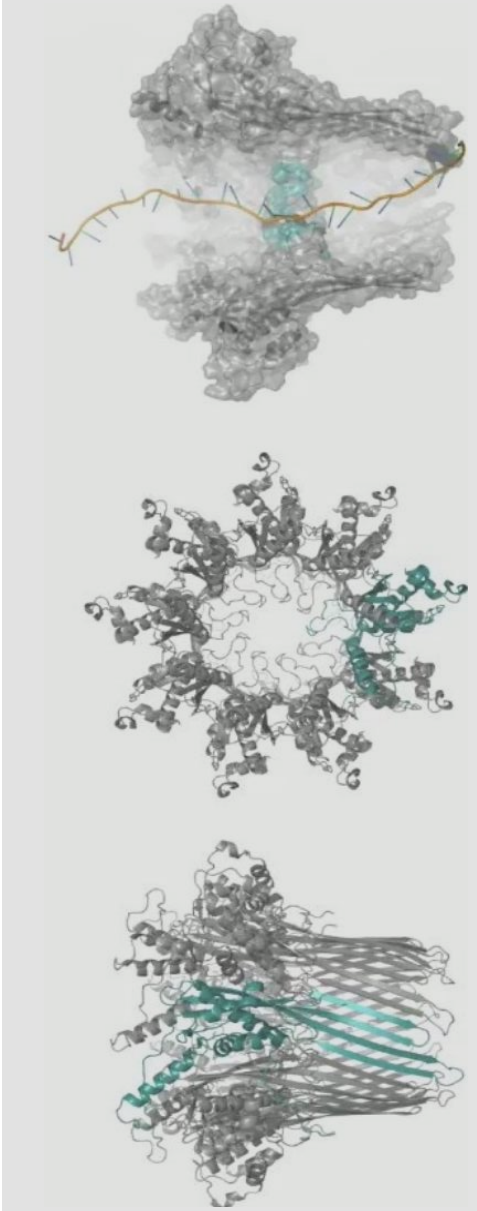
KaruBioNet meeting 13/11/2019

KaruBioNet

Institut Pasteur de la Guadeloupe

TReD-Path
Transmission Réservoirs & Diversité des Pathogènes

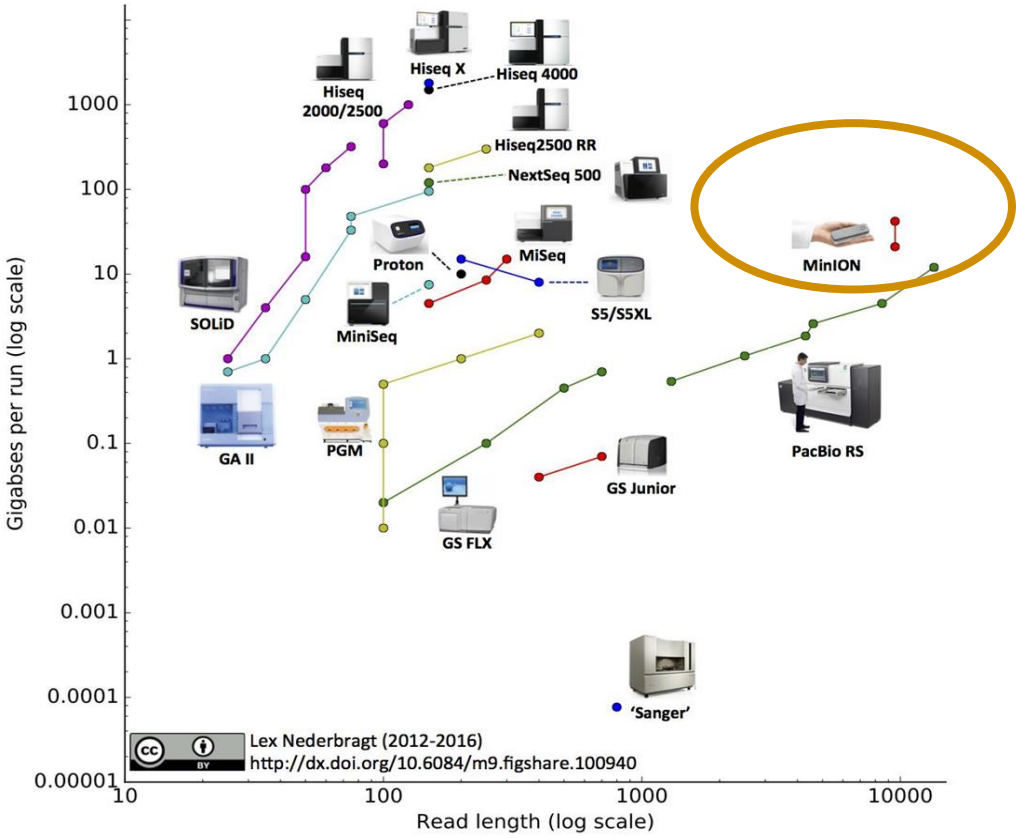
LEMiC



Context

3rd generation sequencing

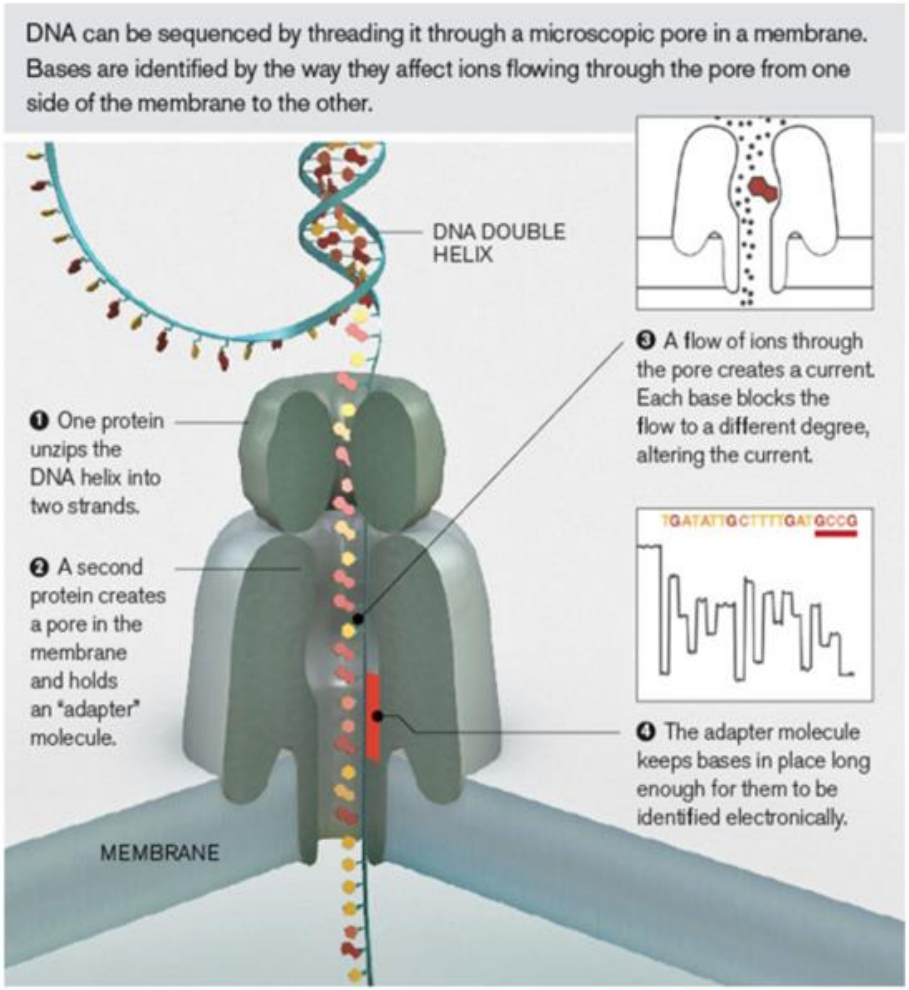
Single-molecule direct sequencing (“third-gen”) vs. sequencing by synthesis (e.g. Illumina)



Institut Pasteur de la Guadeloupe

Context

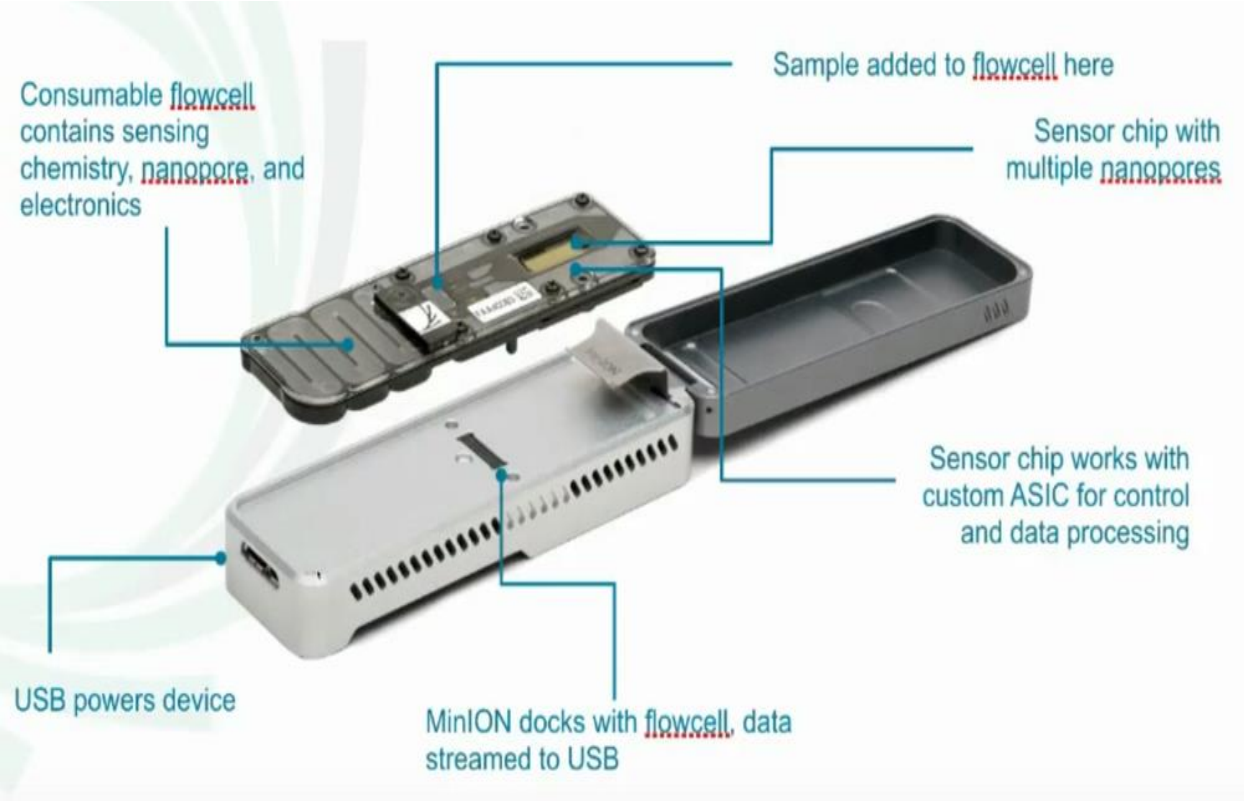
How it works ?



Institut Pasteur
de la Guadeloupe

Context

Anatomy



Institut Pasteur de la Guadeloupe

Context

Fast evolution

2016 2017 2018

Institut Pasteur de la Guadeloupe

Context

Fast evolution

	Flongle	MinION Mk1B	MinION Mk1C	GridION Mk1	PromethION 24
Number of channels per flow cell	126	512	512	512	3000
Number of flow cells per device	1	1	1	5	24
Price per flow cell	$90	$900 - $475	$900 - $475	$900 - $475	$2000 - $625
Run time	1 min - 16 hours	1 min - 48 hours	1 min - 48 hours	1 min - 48 hours	1 min - 72 hours
DNA sequencing yield per flow cell (typical in field - best in field)	1 - 2 Gb	15 - 30 Gb	15 - 30 Gb	15 - 30 Gb	100 - 180 Gb
DNA sequencing yield per device (typical in field - best in field)	1 - 2 Gb	15 - 30 Gb	15 - 30 Gb	75 - 150 Gb	2.4 - 4.3 Tb
Price per Gb, DNA sequencing	**@1 - 2 Gb** $90 per flow cell: 90 - 45	**@15 - 30 Gb** $900 per flow cell: $60 - 30 $790 per flow cell: $53 - 26 $675 per flow cell: $45 - 23 $500 per flow cell: $33 - 16 $475 per flow cell: $31 - 15	**@15 - 30 Gb** $900 per flow cell: $60 - 30 $790 per flow cell: $53 - 26 $675 per flow cell: $45 - 23 $500 per flow cell: $33 - 16 $475 per flow cell: $31 - 15	**@75 - 150 Gb** $900 per flow cell: $60 - 30 $790 per flow cell: $53 - 26 $675 per flow cell: $45 - 23 $500 per flow cell: $33 - 16 $475 per flow cell: $31 - 15	**@2.4 - 4.3 Tb** $2000 per flow cell: $20 - 11 $1700 per flow cell: $18 - 9 $1475 per flow cell: $15 - 8 $1225 per flow cell: $12 - 7 $975 per flow cell: $10 - 5 $625 per flow cell: $6 - 3

Institut Pasteur de la Guadeloupe

Context

Steps

New Technology + Startup Company = putatives Troubleshootings !

1. **Purchasing**
2. **Computer**
3. **Software**
4. **DNA extractions**
5. **Library prep**
6. **Sequencing**
7. **Data analysis**
8. **Data storage**
9. **$ per sample?**

Institut Pasteur
de la Guadeloupe

Steps

1. Purchasing

You need to buy a starter pack before being part of the “Community” ~ 1000 euros

→ 1 minION + 2 flowcells

Institut Pasteur de la Guadeloupe

Steps

2. Computer

Check requirements otherwise no SAV!

Minimum host computer specification

This example will allow a user to run a MinION with real-time local basecalling. Given the amount of data potentially generated, the basecalling may fall behind and need to be completed at the end of an experiment.

Component	Configuration		
Operating Sytem	Windows	Mac	Linux
	7, 8, 10 (64 bit)	Yosemite El Capitain Sierra (64 bit)	Ubuntu 14.4* 16.4
Memory/RAM:	16GB RAM		
CPU	i7 or Xeon**		
Storage	1TB SSD		
Ports	USB3***		

** Linux products are offered under limited support and may take the team longer to respond to queries*

*** Users need to verify their i7 is a four core model or better.*

**** The MinION device is CE marked using USB3. If a user wished to use USBC they may but this invalidates the CE marking*

Institut Pasteur de la Guadeloupe

Steps

3. Softwares

MinKNOW, the operating system

- Data acquisition
- Real time analyses & feeback
- Local basecalling
- Output .fast5 or .fastq files

Guppy contains the ONT basecalling algorithms
(NB other basecaller possible)

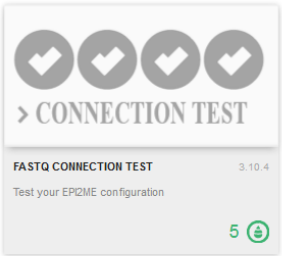
Filtlong
Filer short reads, and decrease randomly reads if too many data generated

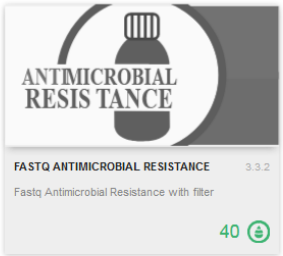
Institut Pasteur
de la Guadeloupe

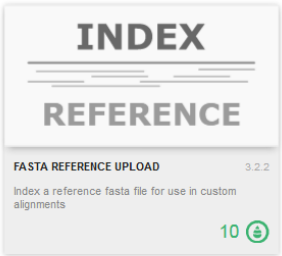
Steps

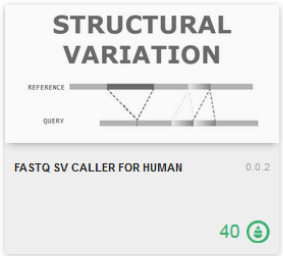
3. Softwares

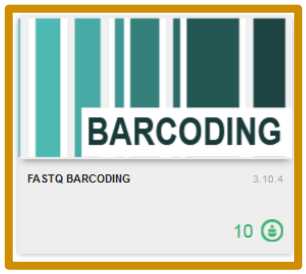
EPI2ME: several workflows



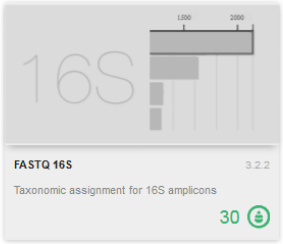


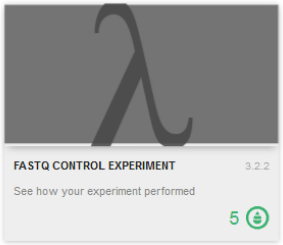


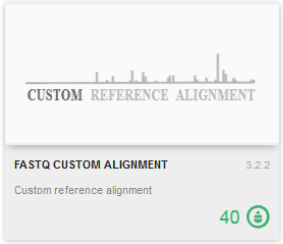


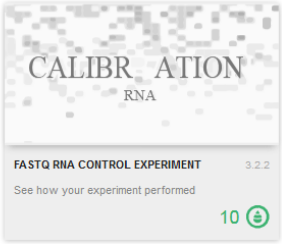


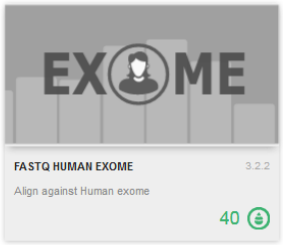
Allow adapter -barcode removal + QC

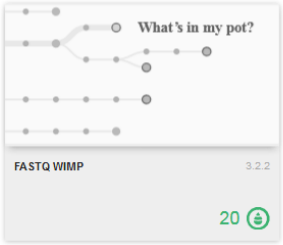












Institut Pasteur de la Guadeloupe

Steps

4. DNA extractions

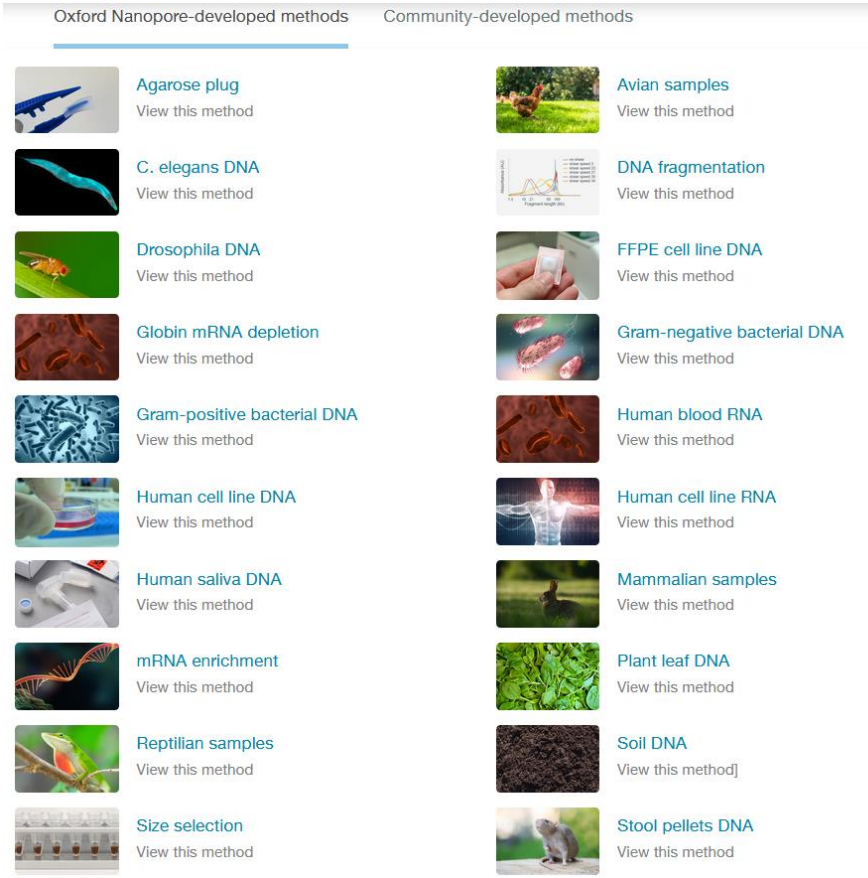
Key step!

- 1 µg needed quantified Qubit
- OD 260/280 1,8
- OD 260/230 2-2,2
- Check integrity on electrophoresis

Some protocols are proposed

But often you have to optimize your own protocol

Ex: hypervirulent klebsiella cultured with salicilate to remove capsule



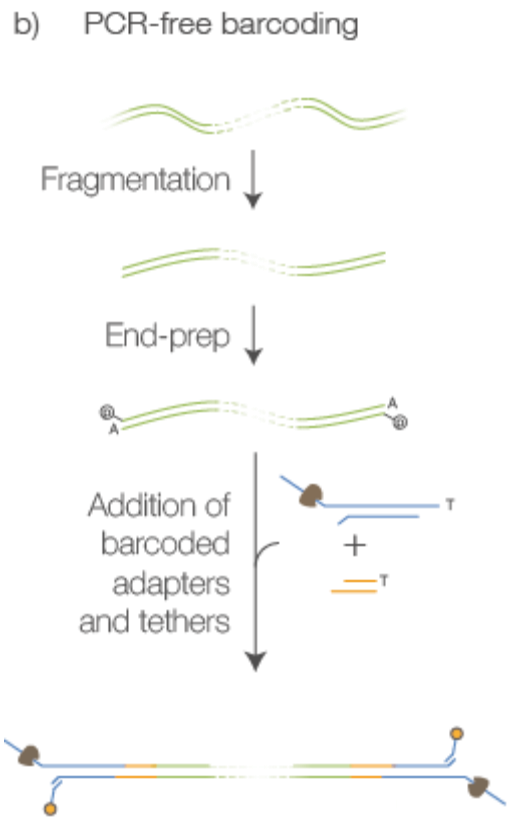
Institut Pasteur de la Guadeloupe

Steps

5. Library prep

Lot of protocols possible:

- we multiplexed n=7 bacterial samples of *E. coli* and *Klebsiella pneumoniae* (24 possible)
- Fragmentation optional
- Kits ONT
    - Native barcoding Expansion 1-12
    - Ligation sequencing kit
    - Flow cell priming kit
- Additional kits need to be ordered:
    - NEBNext FFPERepair mix
    - NEBNext End repair d-A tailing
    - NEB Blunt/TA ligase master mix
    - NEBNext quick ligation module
    - LoBind tubes



Institut Pasteur de la Guadeloupe

Steps

5. Library prep

Basecalled	N°Souche	DO260/280nmGenovaNano	DO260/230nmGenovaNano	Concentration[ng/µl]Qubit	concentration_after_barcode	volume_kept_for_pooling DNA_kept(ng)	ngtotal
BC1	GESC241	1,8	1,4	65,6	46.6	7	326
BC2	GESC242	1,8	1,5	111	38.6	7	270
BC3	GESC249	1,9	1,3	46,4	43.4	7	303
BC4	GESC272	1,9	1,5	63,9	44.8	7	313
BC5	GESC324	1,7	1,2	49,5	55.4	9	498
BC6	SF07	1,3	2,7	3,7	7.74	14	108
BC10	KPB6	1,7	0,5	11,1	14	14	196

sample	DO260/280nmGenovaNano	DO260/230nmGenovaNano	Concentration[ng/µl]Qubit
pooled samples	1,789	1,432	60
volume total loaded: 12 µl. Total DNA 720 ng			

Institut Pasteur de la Guadeloupe

Steps

6. Sequencing & 7. Data analyses

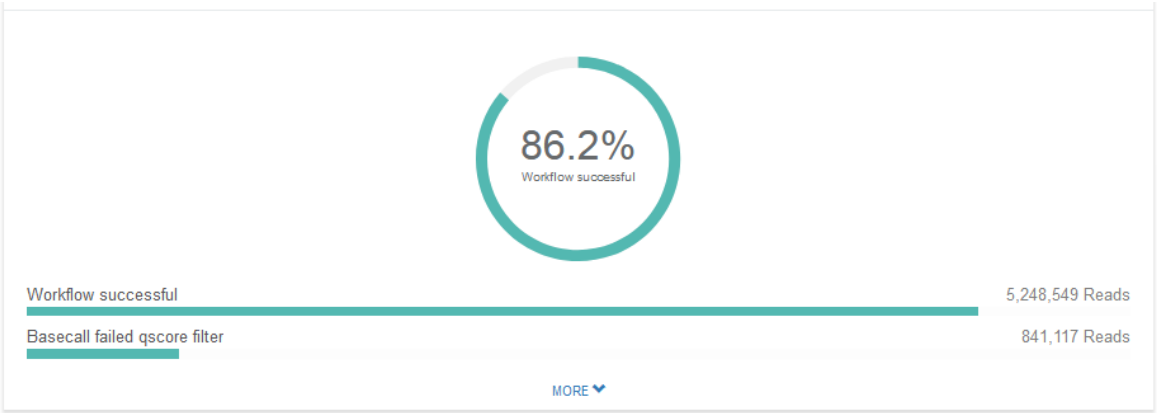
- First check number of pores available (bad if <800): here 1270/2000
- Time determine sequencing depth: can be stopped whenever needed.
  Here 48h
- Only 30% basecalled when the run was stopped
- Basecalling finished on guppy after recovering raw files: very long on classical CPU. So run on UA cluster GPU… 1 day against 1 month!!!

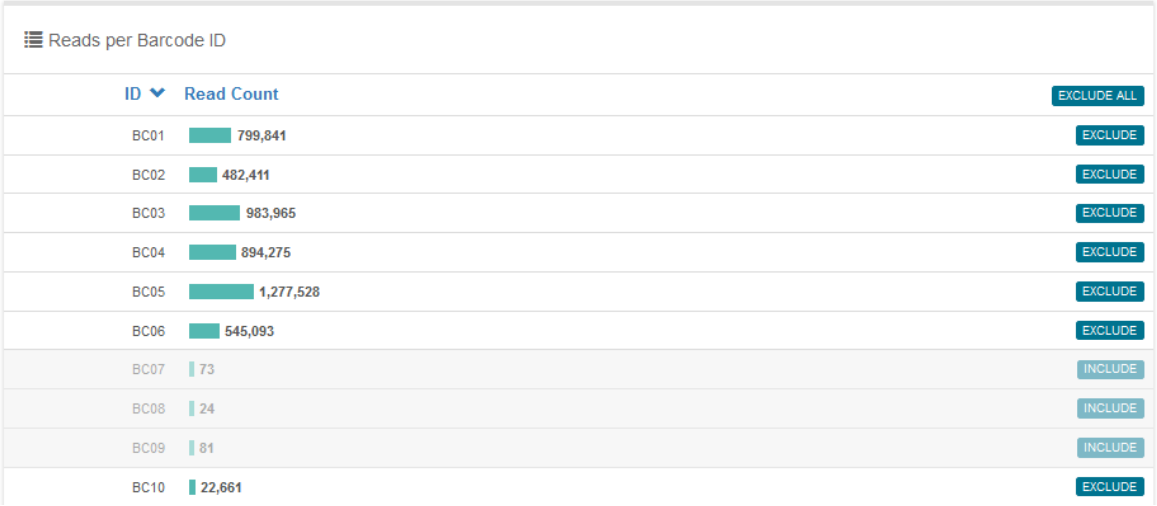
Institut Pasteur
de la Guadeloupe

Steps

7. Data analyses

- EPI2ME for QC, demultiplexing





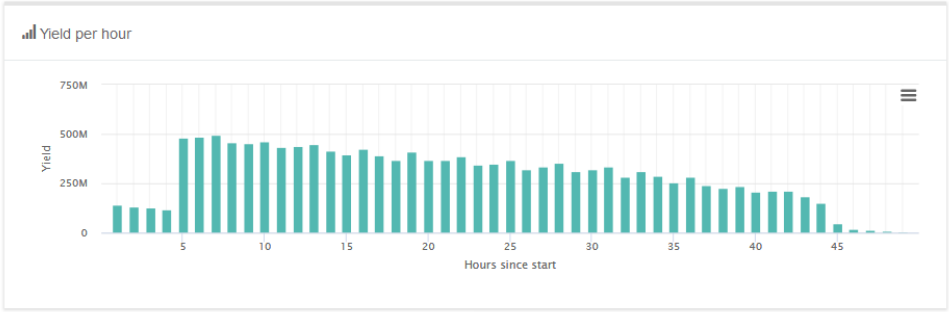
Institut Pasteur de la Guadeloupe

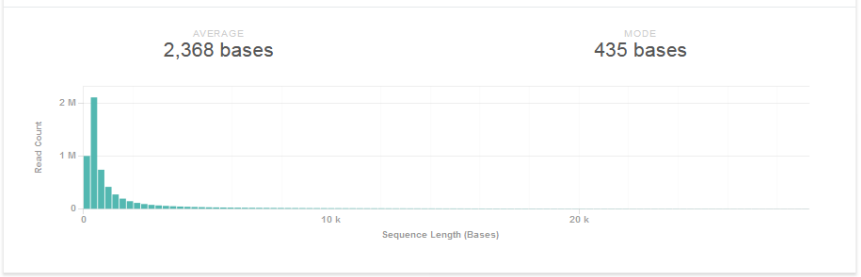
Steps

7. Data analyses

- EPI2ME for QC, demultiplexing
  - Failed for one sample
  - Relatively small fragments
  - But great depth
  - Good quality score

Basecalled	N°Souche	reads_count	~coverage(5Mgenome)
BC1	GESC241	799841	440X
BC2	GESC242	482411	440X
BC3	GESC249	983965	340X
BC4	GESC272	894275	420X
BC5	GESC324	1277528	560X
BC6	SF07	545093	260X
BC10	KPB6	22661	7.92X





Steps

7. Data analyses

Unicycler:

Hybrid assembling between illumina and ONT:
1/ gaps in short read assembly graph are bridged with long read contigs ;
2/ contigs are circularized, overlaps removed and reorientated to dnaA/repA (if possible)
3/ final contigs are polished with short reads data

Handles plasmid-rich genomes
Low misassembly rates
Deal with repetitive genomes
Produce assembly graph for Bandage

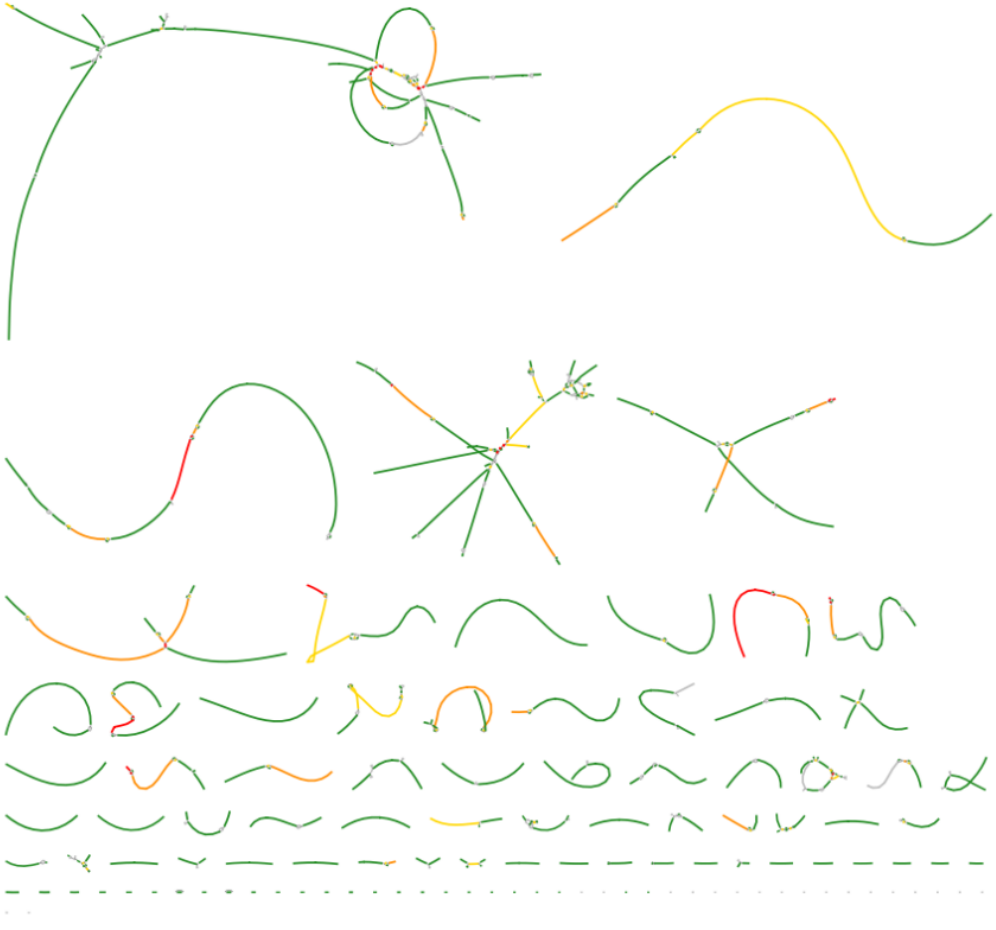
Bandage

Institut Pasteur
de la Guadeloupe

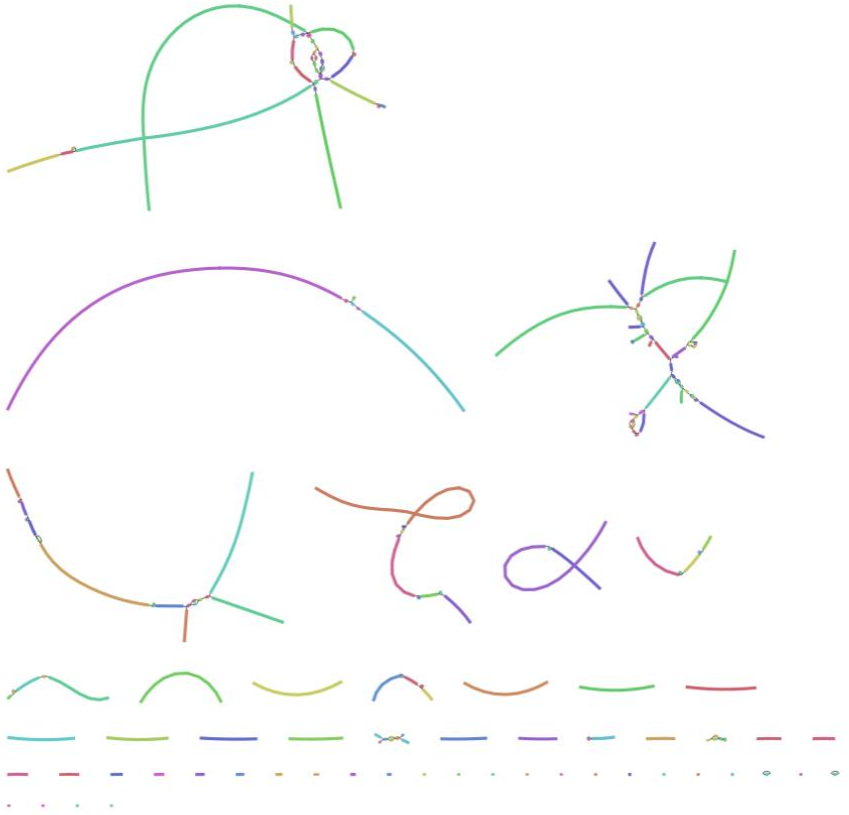
Steps

7. Data analyses

Illumina

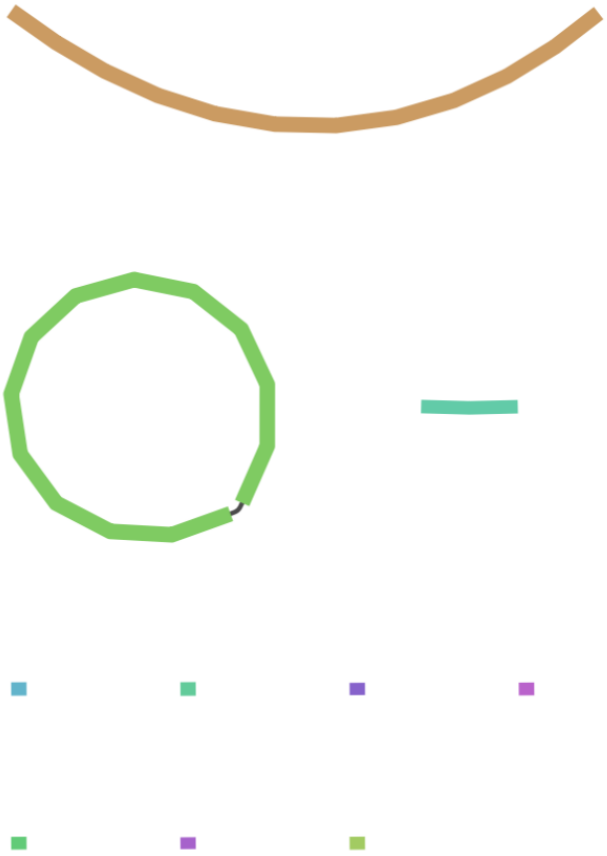


Hybrid Illumina/ONT



Steps

7. Data analyses

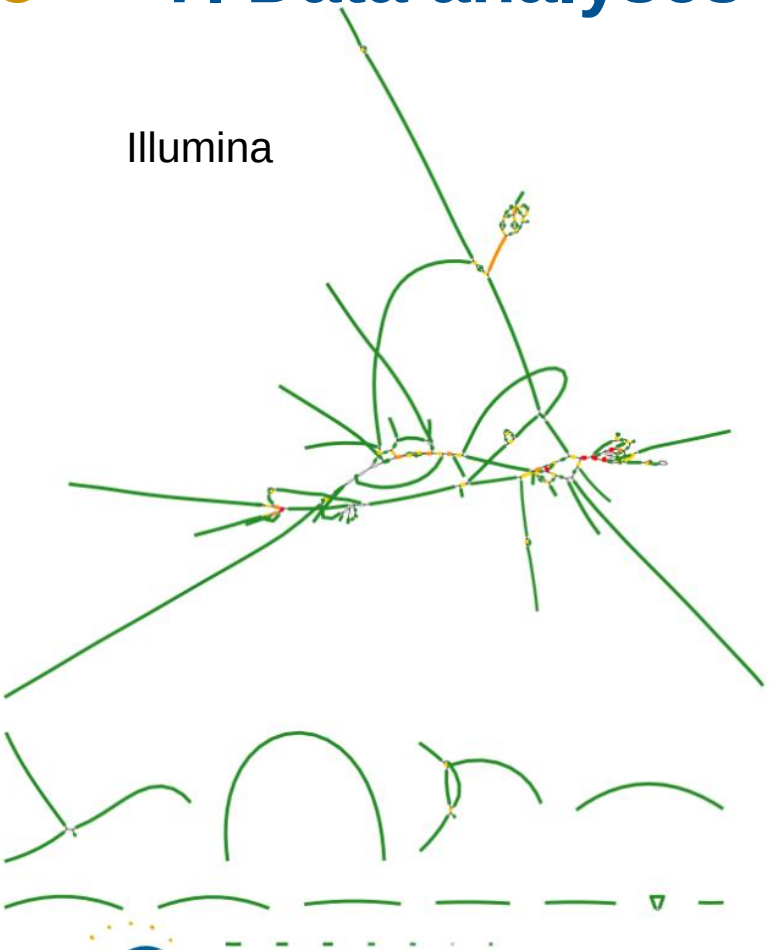


Institut Pasteur de la Guadeloupe

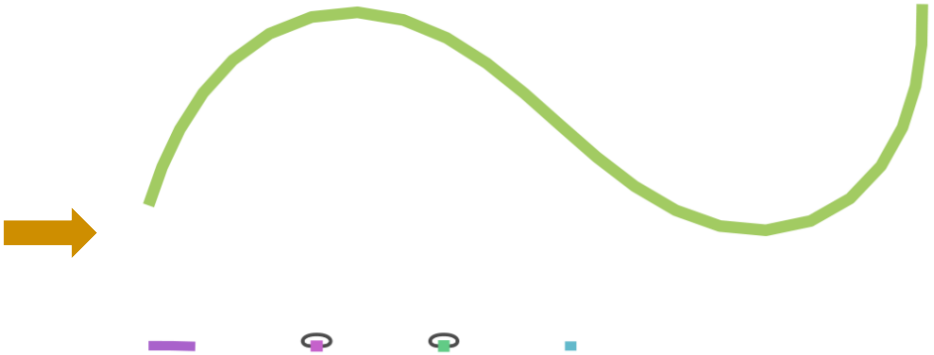
Steps

7. Data analyses

Illumina



Hybrid ONT/Illumina

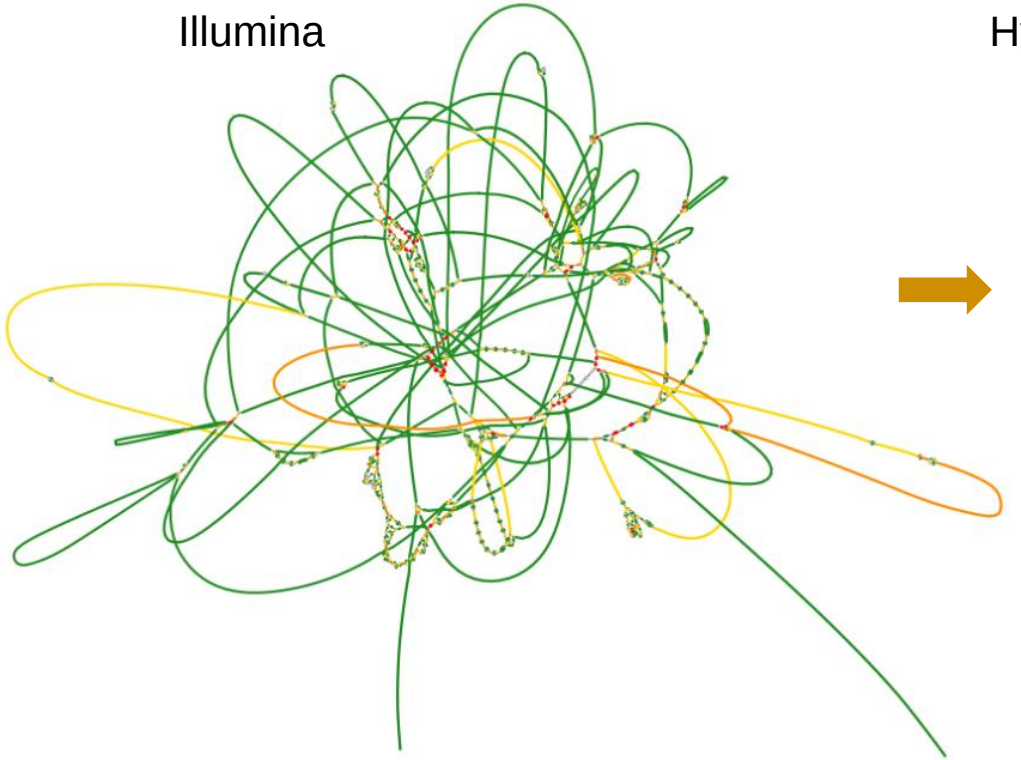


Institut Pasteur de la Guadeloupe

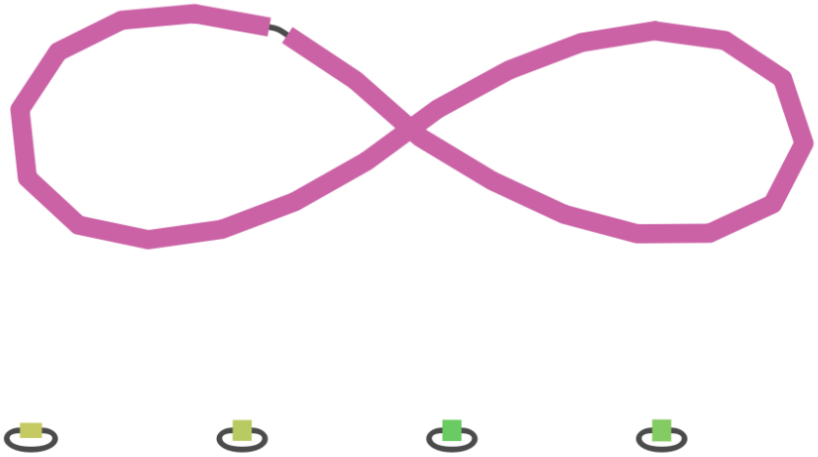
Steps

7. Data analyses

Illumina



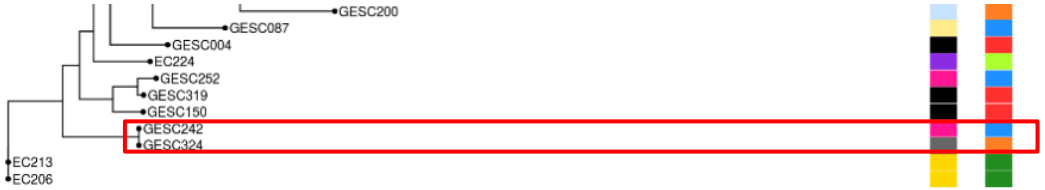
Hybrid ONT/Illumina PERFECT



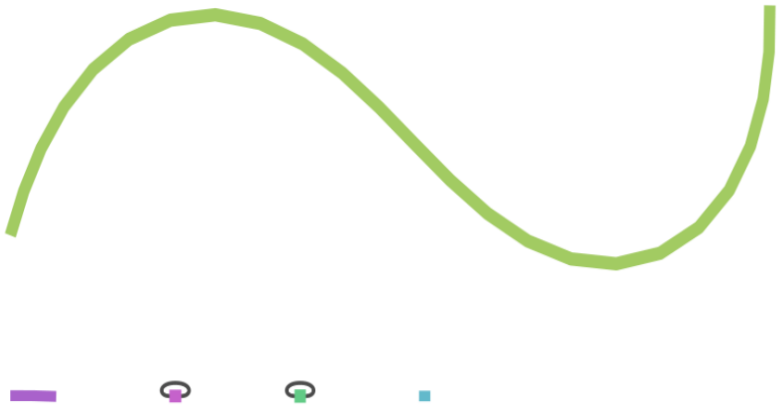
Institut Pasteur de la Guadeloupe

Steps

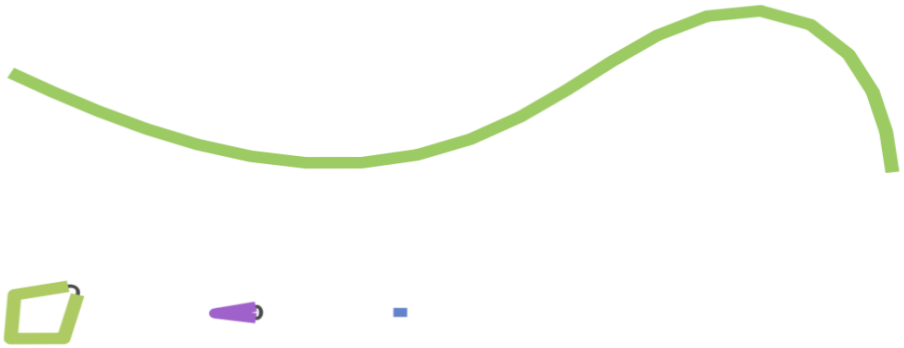
7. Data analyses



GESC242 (oiseau, STEP)



GESC324 (cafard, égout CHU)



#GESC242

Component	Segments	Links	Length	N50	Longest segment	
total	5	2	5,040,336	4,745,166	4,745,166	
1	1	0	4,745,166	4,745,166	4,745,166	incomplete
2	1	0	204,105	204,105	204,105	incomplete
3	1	1	53,644	53,644	53,644	complete
4	1	1	34,653	34,653	34,653	complete
5	1	0	2,768	2,768	2,768	incomplete

GESC324

Component	Segments	Links	Length	N50	Longest segment	Status
total	4	2	5,133,535	4,012,746	4,012,746	
1	1	0	4,012,746	4,012,746	4,012,746	incomplete
2	1	1	803,514	803,514	803,514	complete
3	1	1	309,211	309,211	309,211	complete
4	1	0	8,064	8,064	8,064	incomplete

Institut Pasteur
de la Guadeloupe

Steps

7. Data analyses

Resfinder database

GESC242

blaCTX-M-15
blaOXA-1
aac(6')-Ib-cr
tet(A)
aac(3)-Iia
aac(6')-Ib-cr
dfrA14
catB3

204,105

blaTEM-1B
qnrB1
qnrS1
aph(3'')-Ib
aph(6)-Id
dfrA14

53,644

IncN

GESC324

probably chromosomal?
Chro 4M
Contig 2 800 kb

blaCTX-M-15
blaOXA-1
blaTEM-1B
aac(6')-Ib-cr
qnrB1
tet(A)
aac(3)-IIa
aac(6')-Ib-cr
aph(3'')-Ib
aph(6)-Id
dfrA14
catB3
sul2

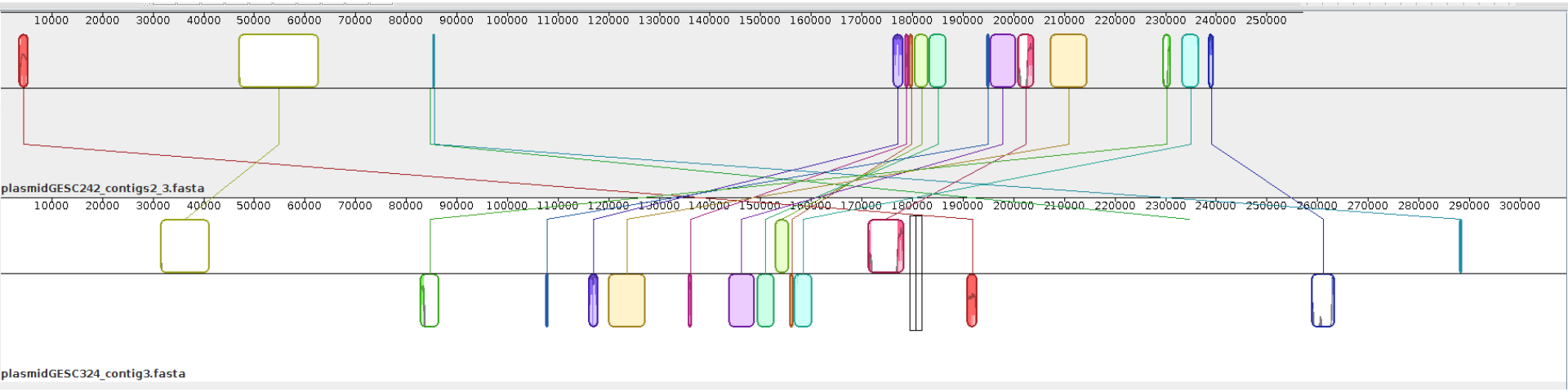
IncH12
309,211

Institut Pasteur de la Guadeloupe

Steps

7. Data analyses

Mauve syntheny: seems to be different plasmids sharing close ATBR profiles
OR
Assembly with error due to repetitive elements unresolved by ONT (to short read ~2500 kb



Institut Pasteur de la Guadeloupe

Steps

8. Data storage

~ 5 Go / ONT .fastq file (for bacteria)

Take care to remove all files when finished

If you perfom live basecalling and if you have less than 1To available, MinKNOW might crash!

Institut Pasteur de la Guadeloupe

Steps

9. Cost

Computer 1To SSD, i7, 16 Go ram= 1400 euros

Barcoded *12 bacteria <100 euros/strain

Putatively 50 euros if barcoded *24! (~ Illumina)

Institut Pasteur
de la Guadeloupe

Conclusion

- gDNA extraction needs to be optimized
  - Increase size of fragments to resolve repetitive elements
  - Quality of DNA should be improved 260/230
  - Protocol ajusted for Kp hypervirulent
- Basecalling done on Guppy at UA instead of live basecalling
- Barcoding *24 should be tested
- PFGE to check that we sequence correct size of plasmids?

Institut Pasteur
de la Guadeloupe