

MinION, GridION, how does Nanopore technology meet the needs of our users ?

Journée Long Reads GeT
28 Novembre 2017

Catherine Zanchetta & Maxime Manno



Wet Lab

- Nanopore technology
- Library preparation
- What can we expect ?

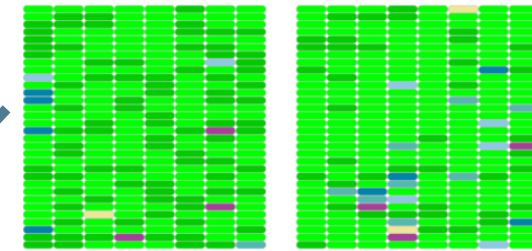


Bioinformatics

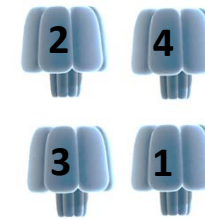
- IT and Bio-informatics solutions
- Some results about Nanopore sequencing and assembly

Nanopore technology

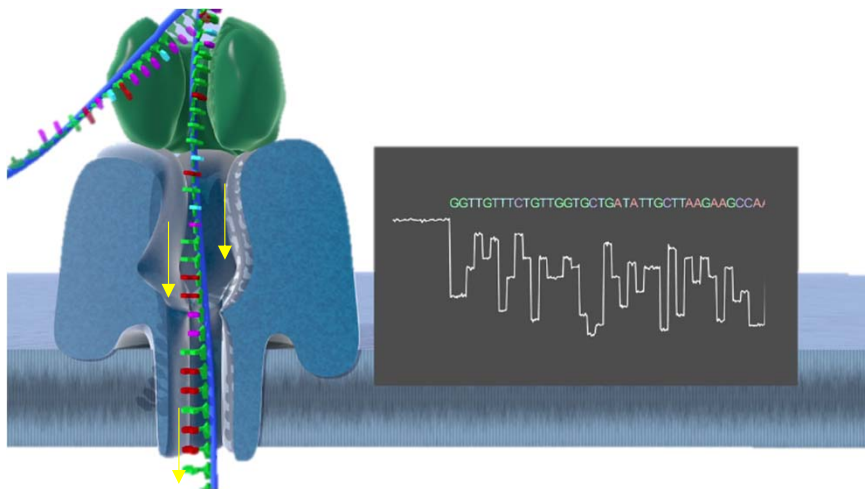
How it works



512 channels



4 pores per channel



- A **protein** is set in an electrically resistant polymer membrane
- An **ionic current** is passed through the nanopore
- The **event / base** creates a characteristic disruption in current
- Identification of G, A, T and C bases

Nanopore technology

A scalable technology

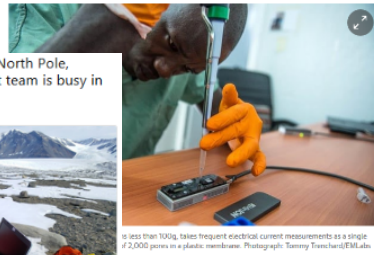
From Ebola to Zika, tiny mobile lab gives real-time DNA data on outbreaks

A genomic surveillance system which fits in a suitcase can help health workers to quickly understand the spread of viruses and break the chain of infection

While the elves are busy at North Pole, Extreme Microbiome Project team is busy in Antarctica.



08.03 - 23.04. 2018
41 followers 80 likes



MinION

~ 7Gb / 48h

2016



GridION

~ 35Gb / 48h

2017



PromethION

~ 2Tb / 48h ???

2018

Nanopore technology

MinION and GridION



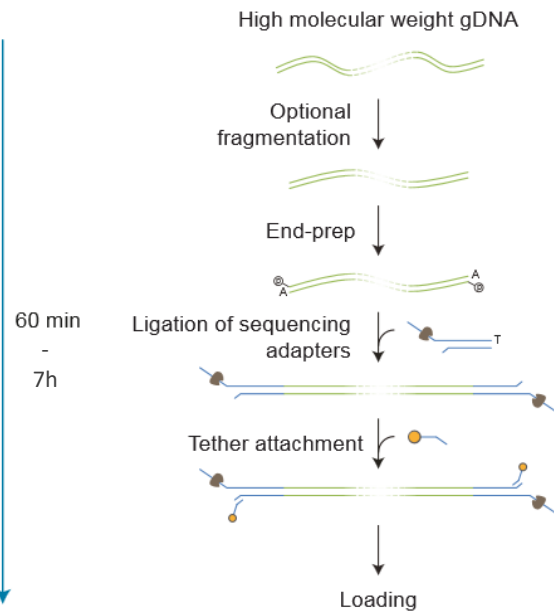
Validated last summer
Same quality and quantity of data in
comparison with the MinION

Number of FC	1	5
Live basecalling	We don't use it	We use it !
Basecaller version	Last version available	Delay for the incorporation of the last version into dogfish
Price per FC	675€	300€

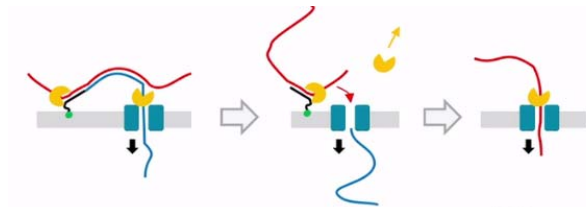
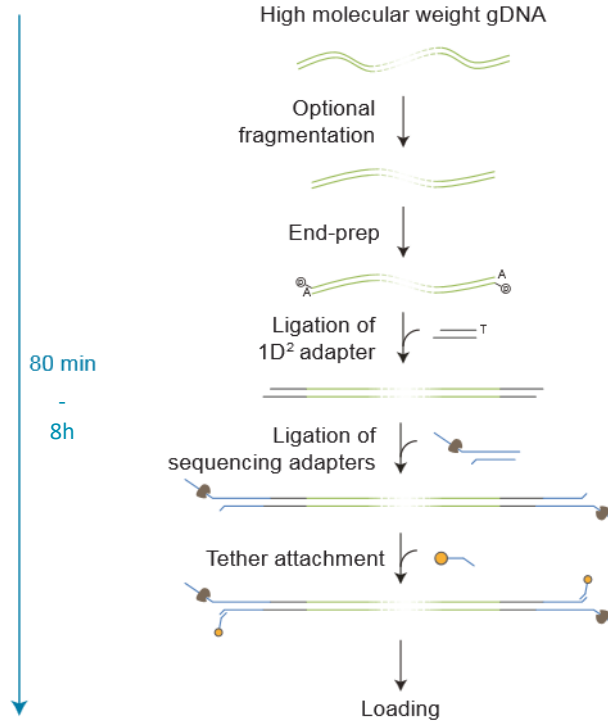
Library preparation

1D and 1D² kits

1D



1D²



Library preparation

Importance of the quality controls

- Have a **good quality DNA**



Nanodrop

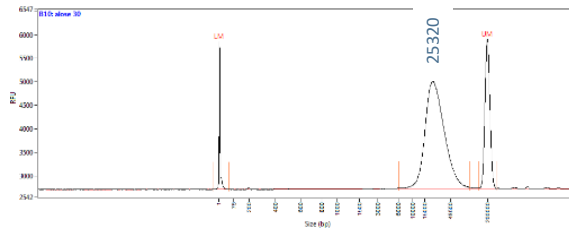
$A_{260}/A_{280} = 1.8-2.2$

$A_{260}/A_{230} = 1.8-2.2$

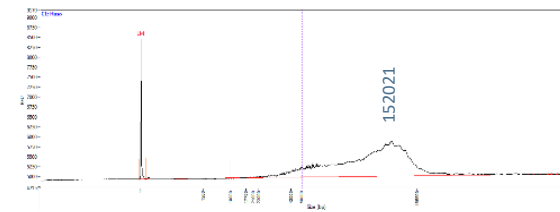
- Start the library prep with the **right number of molecules**



Qubit



Fragment analyzer
DNA < 40 kb



FEMTO
DNA > 40 kb



Library preparation

Suitable library prep

- Make a **suitable library prep**

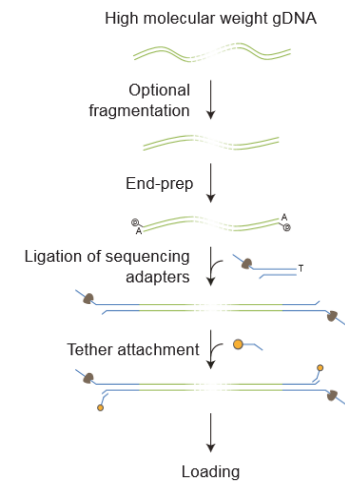
Megaruptor shearing
(improves the yield ; if necessary)



BluePippin
(Removes small fragments ; if necessary)



1D lib prep
(optimized for long fragments)



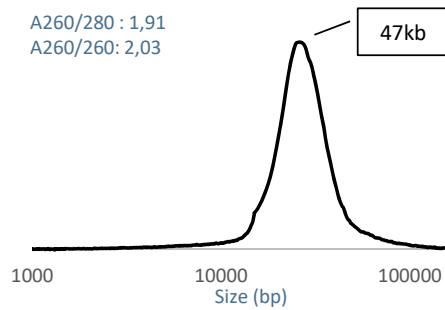
1.8 μ g / Gb of data

What can we expect ?

The read distribution depends on the DNA quality

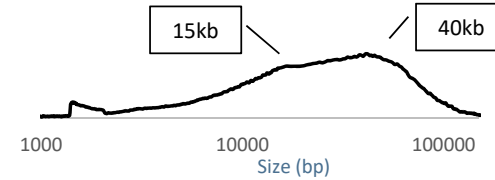
Non degraded DNA

A260/280 : 1,91
A260/260 : 2,03



Degraded DNA

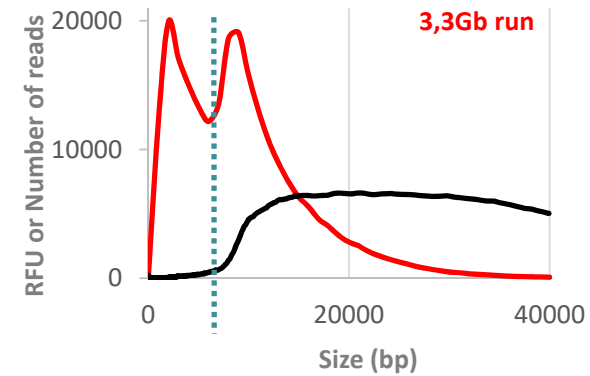
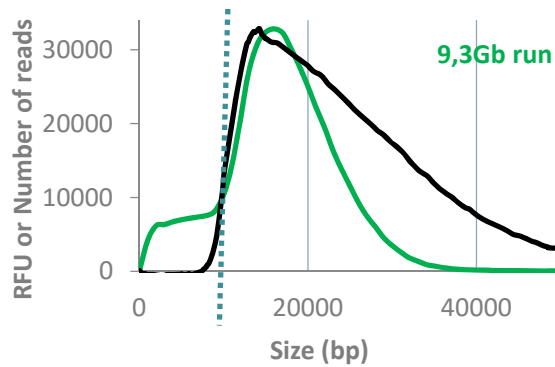
A260/280 : 1,94
A260/260 : 1,97



20kb shearing



>10kb selection



Library

Reads

Blue Pippin Cutoff

What can we expect ?

The yield might be very different from a flowcell to another



15-30 kb



Only if necessary



Similar number of molecules loaded

Gb/FC

Good quality DNAs

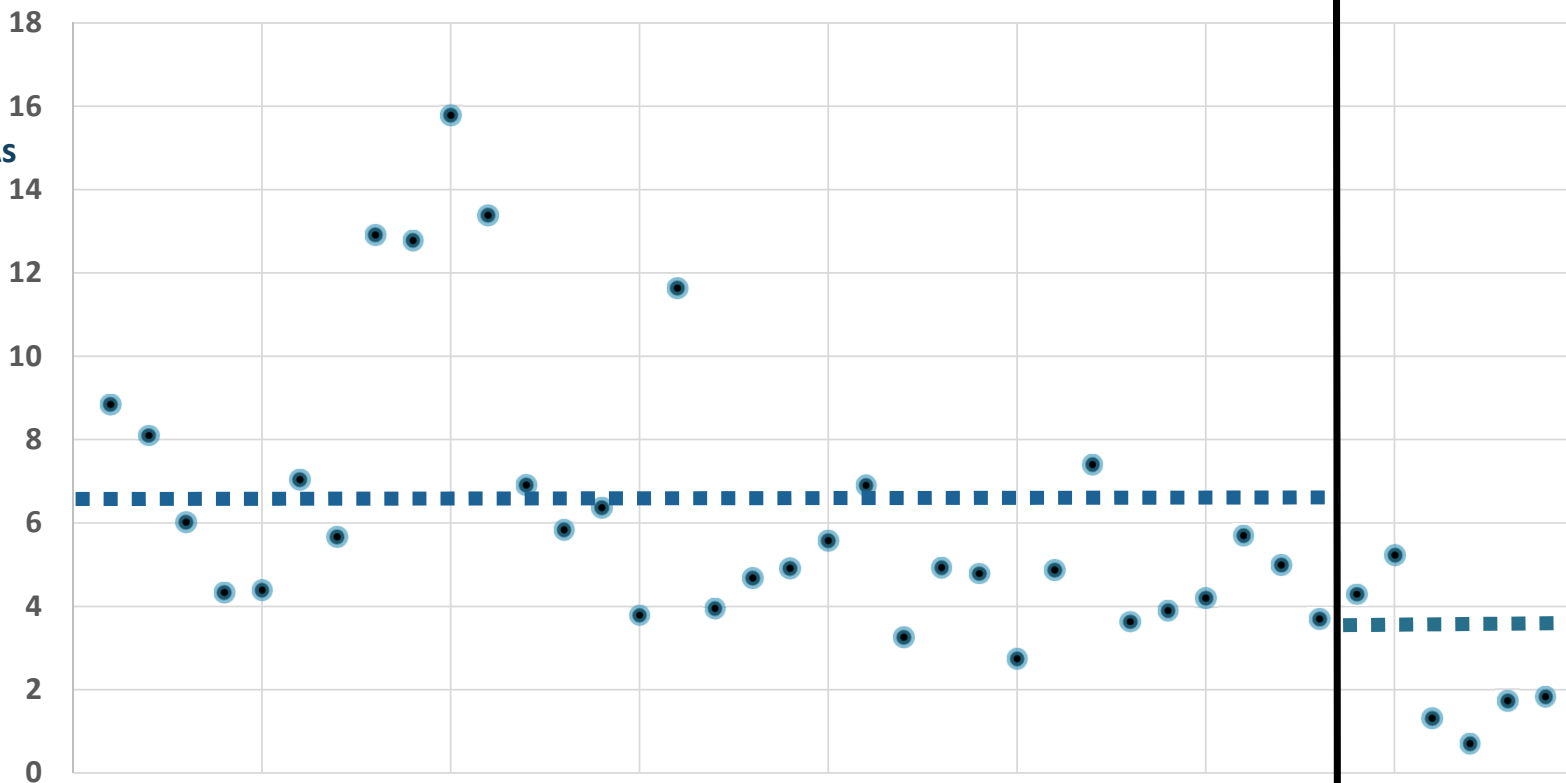
A260/280 ; A260/230
1.8-2.2

Mean yield : 6.5 Gb
~ 70 % reads >Q10

Bad quality DNAs

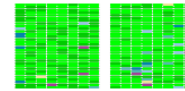
A260/280 ; A260/230
<>1.8-2.2

Mean yield : 2.5 Gb
~ 40 % reads >Q10



What can we expect ?

15kb – 20kb DNA : preferred size

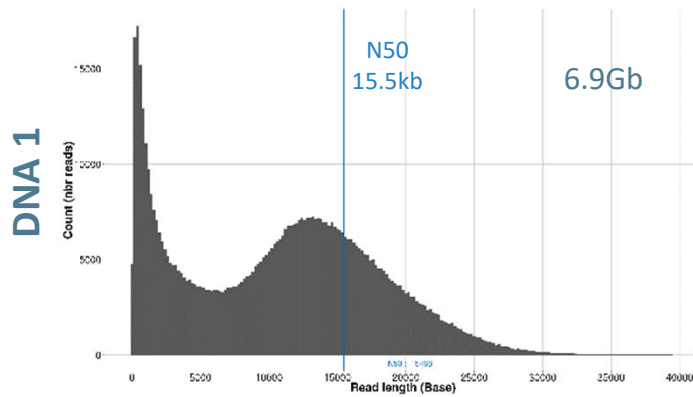


Similar number of active pores

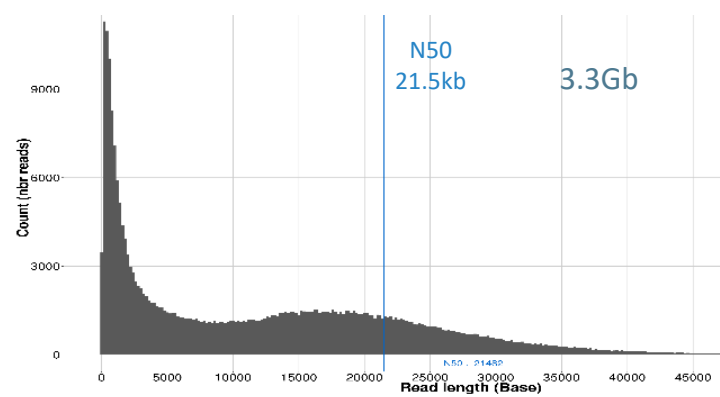


Similar number of molecules loaded

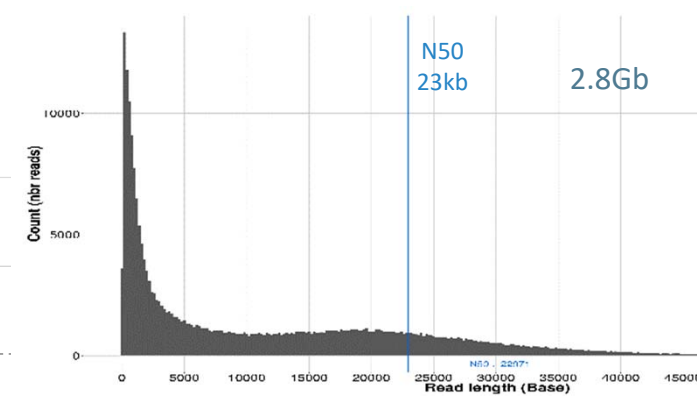
20kb shearing



30kb shearing

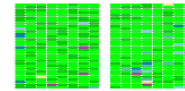


40kb shearing



What can we expect ?

15kb – 20kb DNA : preferred size



Similar number of active pores



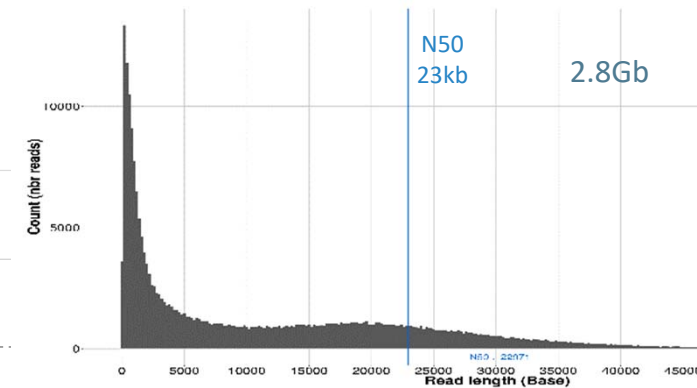
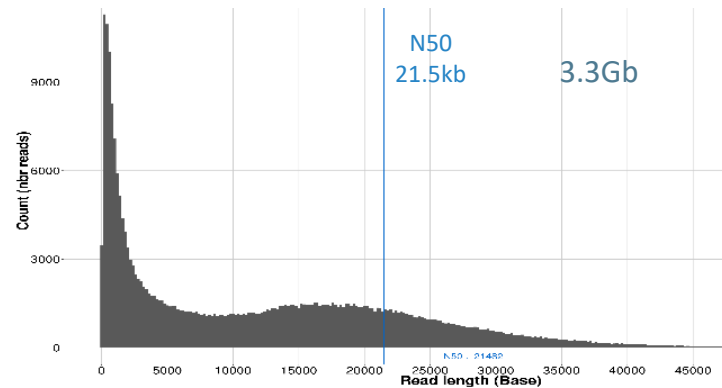
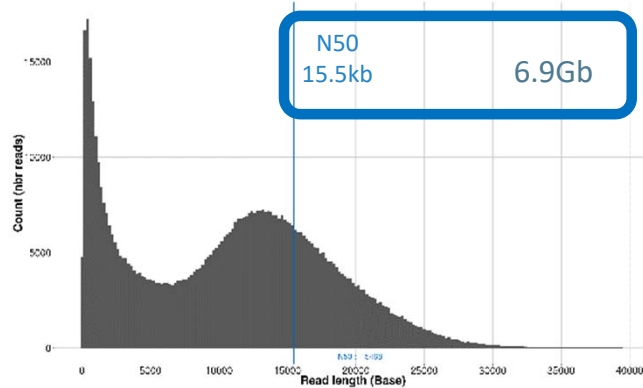
Similar number of molecules loaded

20kb shearing

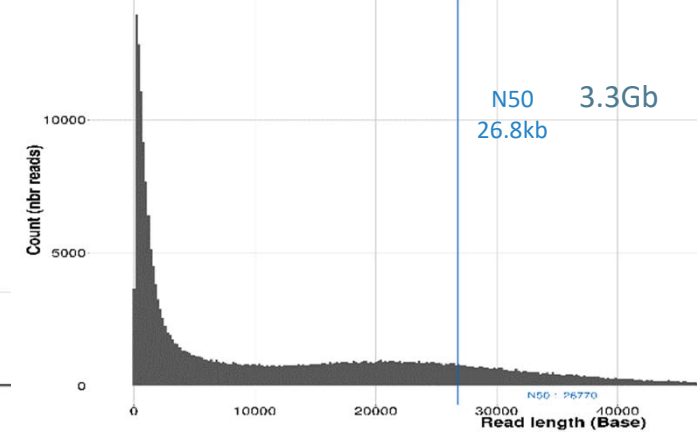
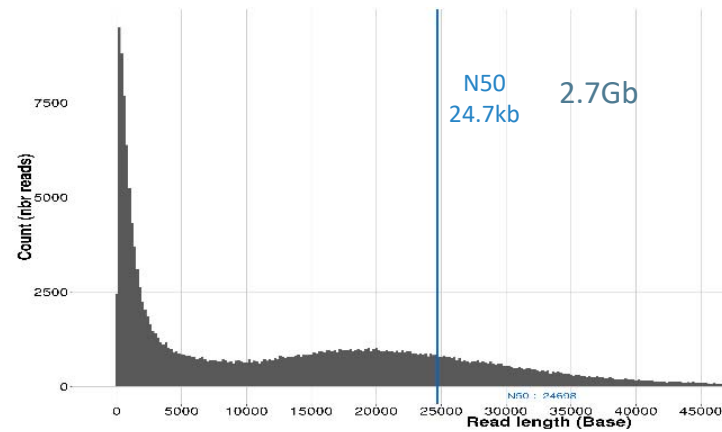
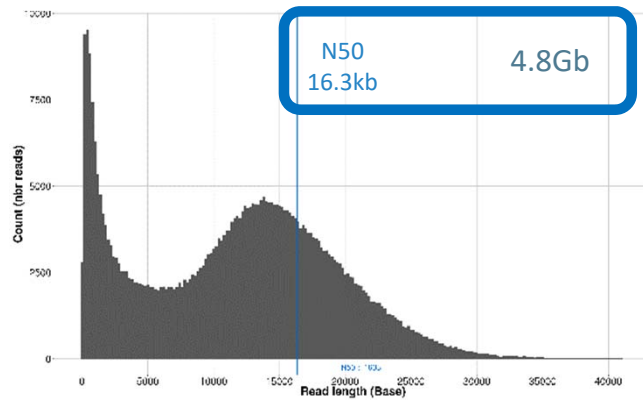
30kb shearing

40kb shearing

DNA 1

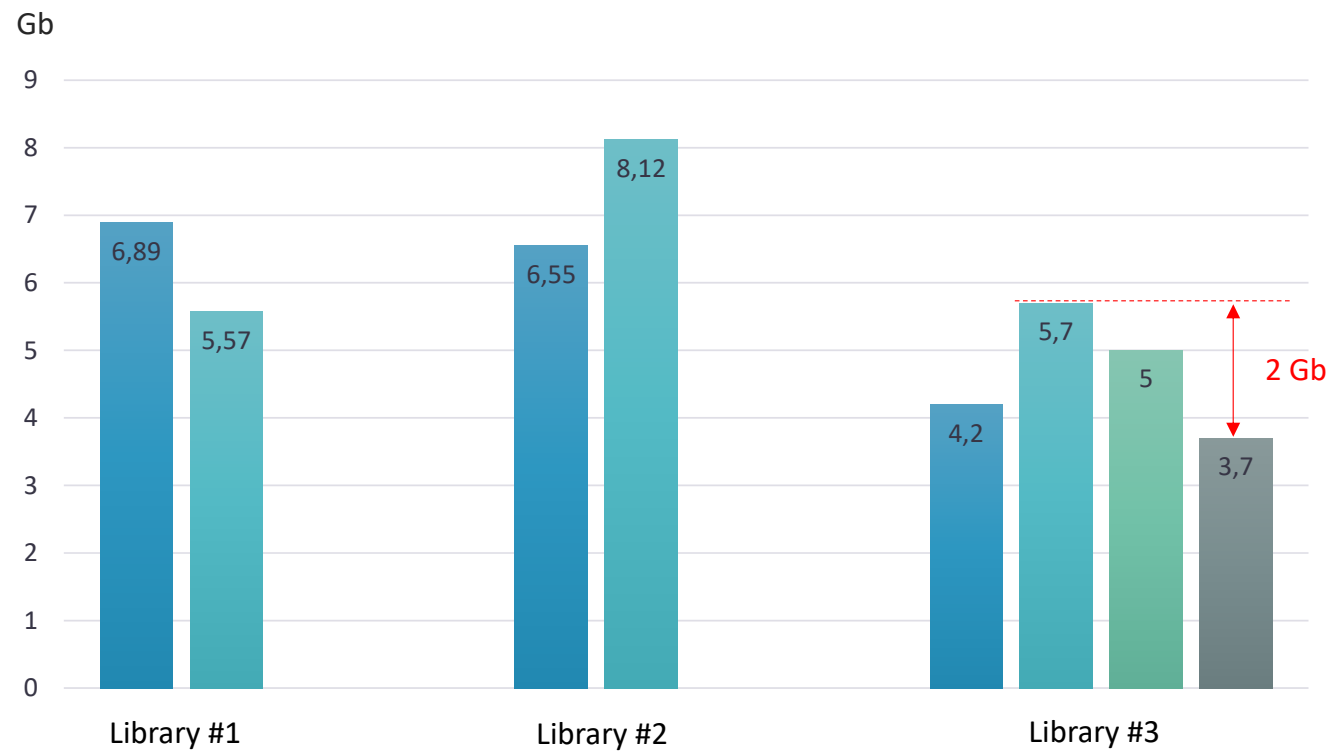
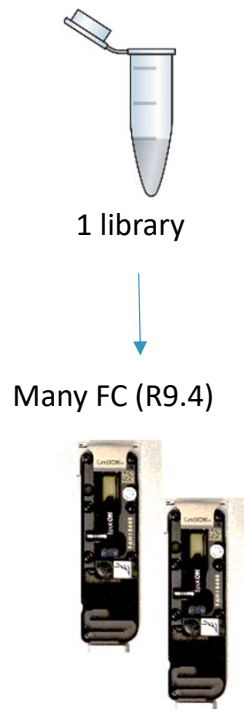


DNA 2



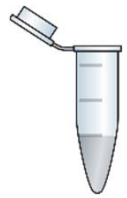
What can we expect ?

Yield variability between flowcells



What can we expect ?

Yield variability between flowcells



1 library

Many FC (R9.4)



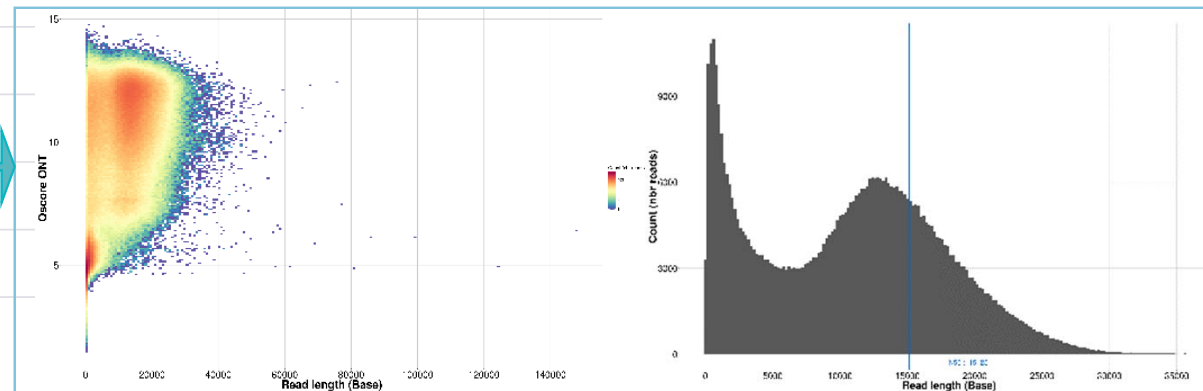
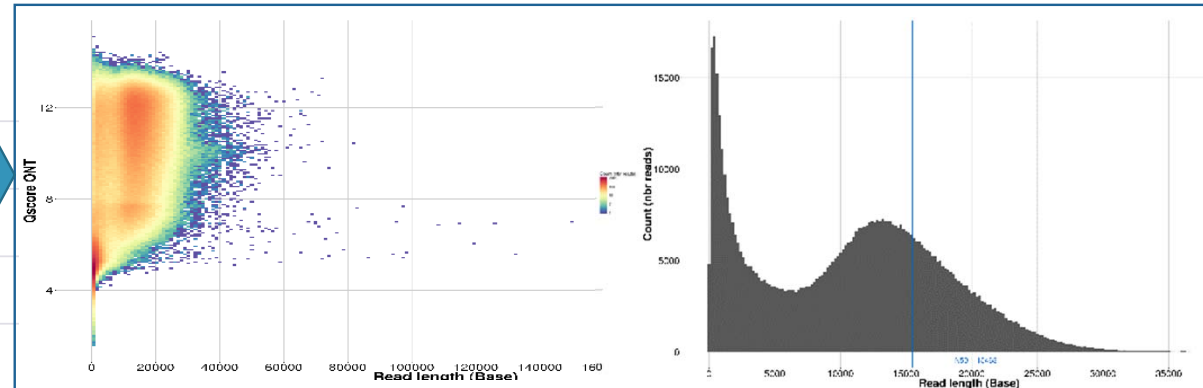
Gb

9
8
7
6
5
4
3
2
1
0

6,89

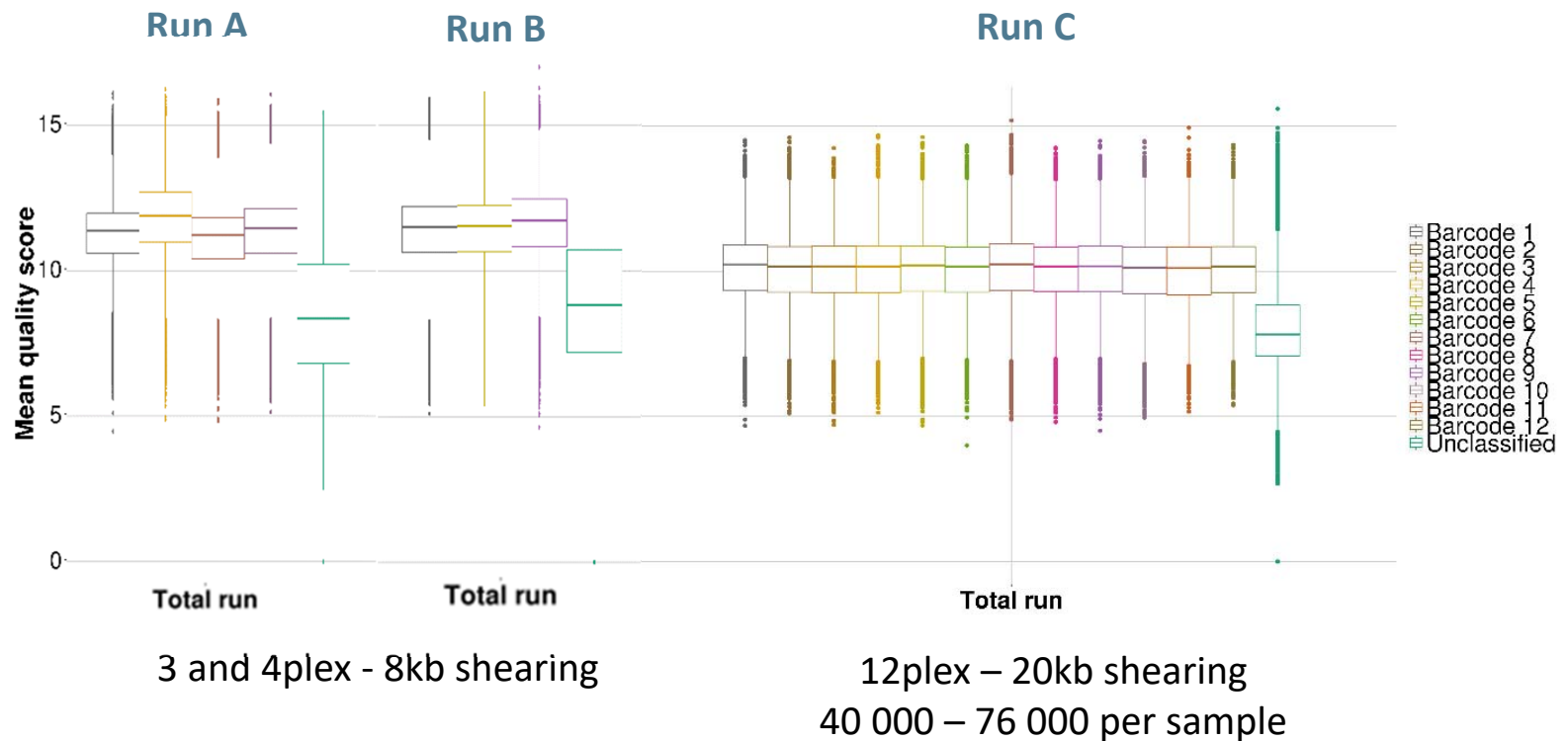
5,57

Library #1



What can we expect ?

Multisample run to lower the cost



Similar quality between samples and runs
More than 65% of reads assigned to a barcode

A et B : R9.4 FC / 3Gb runs / Minknow 1.5 / Albacore 1 ; C : R9.4 FC / 13.4Gb run / Minknow 1.7.3 / Albacore 1.1.0

What can we expect ?

Longest read : 1Mb read

Rapid protocol



Martin A. Smith
@martinalexsmith

Suivre

970kb @nanopore read maps contiguously to 1,033,955nt of chr19. Still have 33h of sequencing to go on this run.

À l'origine en anglais

05:48 - 27 oct. 2017

51 Retweets 135 J'aime



3

51

135



Extraction : key step



Wet Lab

- Nanopore technology
- Library preparation
- What can we expect ?



Bioinformatics

- IT and Bio-informatic solutions
- Some results about Nanopore sequencing and assembly



Qscore ONT

ONT data

Dogfish

Minknow

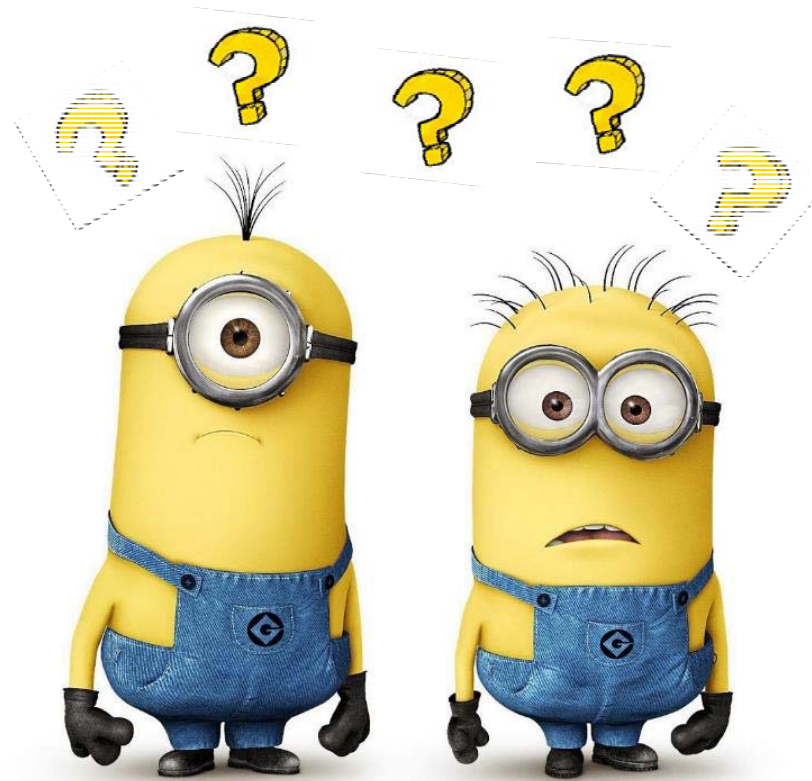
Porechop

MinION

Albacore

Polishing

Fast5 file



Easy ???

What is it ? How to do it ?

Thanks to :

Strong collaboration with the Genotoul Bio-informatics Plateform.

- IT resources
- Great help





- ⑤ **IT and Bio-informatic solution**
 - Sequencing and raw data (MinION, GridION)
 - Trimming and filtering
 - Assembling and polishing
- ⑤ **Some results about Nanopore sequencing and assembly**
 - Deadlines and computing time
 - Raw data : Error rate and Qscore
 - Assembly : Completeness

IT solution : Sequencing



Sequencing Minknow
C:\data\reads\
Fast5 files = raw data

Live BaseCalling :
Dogfish and Albacore
Fast5 = raw data + base
sequence and Fastq

Transfer from
GridION to
Genotoul server

Transfer from
MinION Computer
to **Genotoul server**

BaseCalling Albacore v2.0
Fastq only

Location :
MinION Computer/ GridION
Genotoul Server

Bio-informatic solution : Trimming and Filtering

TRIMMING – Porechop

→ Output : Trimmed fastq

Finding and removing Nanopore adapters

67.4% of reads had adapters trimmed from their start (0.01% bp removed)

20% of reads had adapters trimmed from their end (0.002% bp removed)

Porechop



<https://github.com/rrwick/Porechop>



Filtlong

FILTERING – Homemade tool (or Filtlong)

→ Output : filtered reads

Size > 3 kb

Qscore > 10

<https://github.com/rrwick/Filtlong>

Cleaned
FASTQ

Bio-informatic solution : Assembly and polishing

Bacterial genome (5Mb)

<https://github.com/marbl/canu>

ASSEMBLY – CANU

→ Output : Contigs

~100X coverage

(POLISHING – Racon x2)

Use Nanopore rawdata to
correct the assembly

<https://github.com/isovic/racon>

POLISHING – Pilon (or Nanopolish)

30X Illumina coverage for Pilon



<https://github.com/broadinstitute/pilon/wiki>

Cleaned
Contig

Cleaned
Fastq



⑤ IT and Bio-informatic solution

- Sequencing and raw data (MinION, GridION)
- Trimming and filtering
- Assembling and polishing

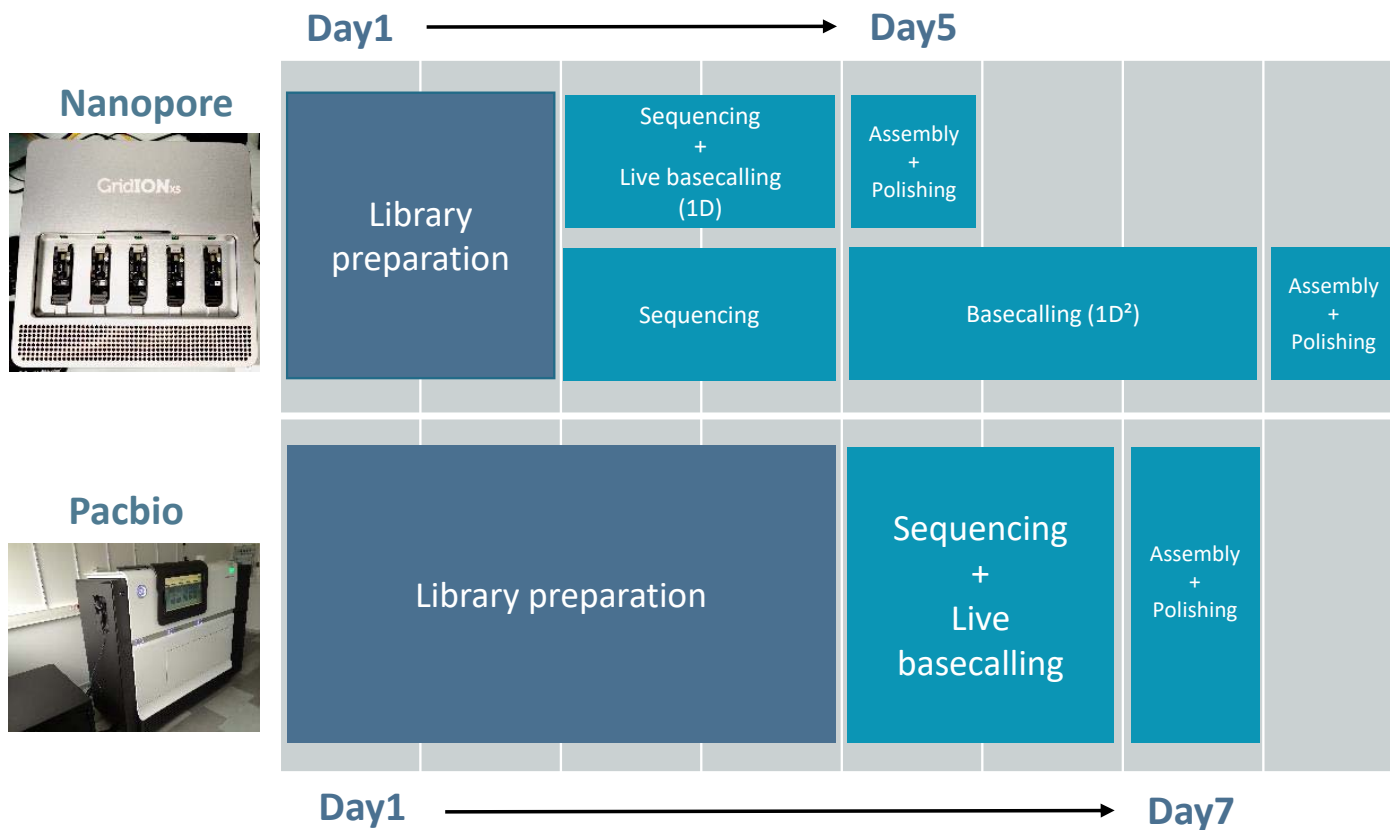
⑤ **Some results about bacterial genome sequencing**

- **Deadlines and computing time**
- **Raw data : Error rate and Qscore**
- **Assembly : Completeness**

Meet the deadlines

How long does it take to get the results ?

Example of a bacterial genome assembly :
Same quantity of data / Same informatics resources



Raw data : Pacbio vs Nanopore

Error rate

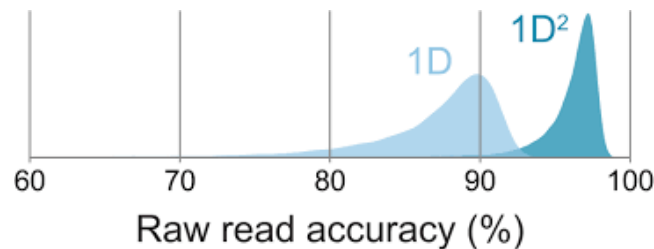
Raw read error rate

Example of a bacterial genome assembly :
Same quantity of data / Same informatics resources

Raw data sets	Error rate ¹
HiSeq (Illumina total raw data)	0.3 %
RSII (PacBio total filtered subreads)	18.7 %
MinION 1D (ONT filtered reads)	12.2 %
MinION 1D ² (ONT filtered reads)	8.8 %

¹Error rate base on alignment to the PacBio genome reference (bwa mem)

Oxford Nanopore announcement



Qscore ? (Quality score for ONT data)

Raw data : Error rate per Qscore

Example of a bacterial genome assembly :
Same quantity of data / Same informatics resources

Run 1D R9.4 MinION 20Kb

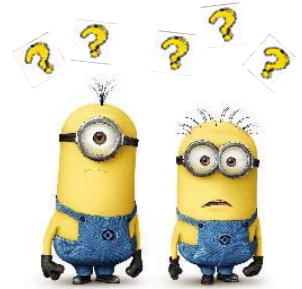
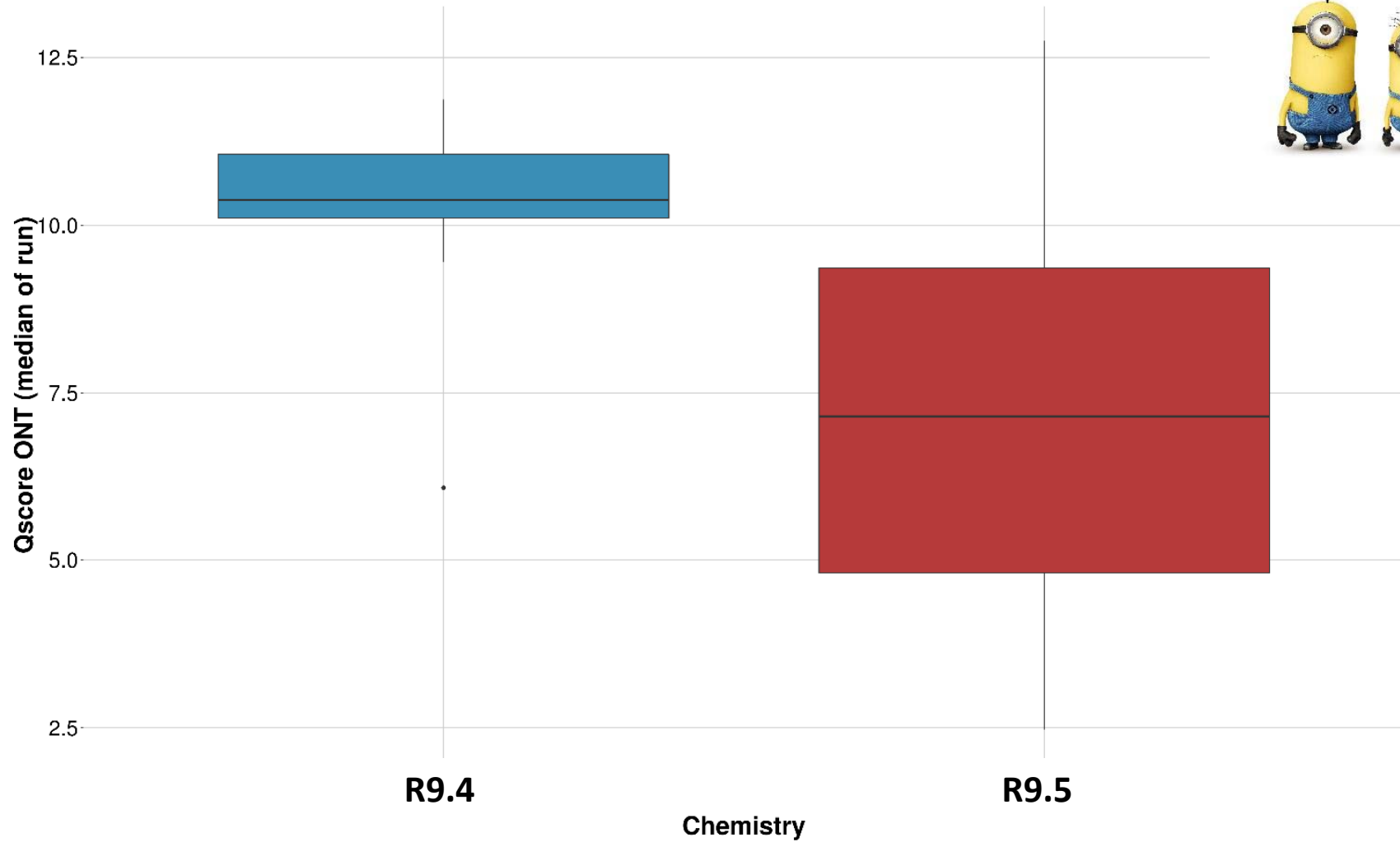
Subset	%reads	Mean Qscore	Error rate
0>Qscore>5	0,1%	4,75	45,7%
5>Qscore>8	11,6%	7,05	34,4%
8>Qscore>10	17,7%	9,13	23,7%
10>Qscore>12	42,3%	11,14	15,1%
12>Qscore>15	28,3%	12,44	10,7%
15>Qscore>20	0,0%	-	-

Run 1D R9.5 GridION 20Kb

Subset	%reads	Mean Qscore	Error rate
0>Qscore>5	3,0%	4,42	44,3%
5>Qscore>8	21,6%	6,67	30,8%
8>Qscore>10	29,7%	9,07	19,1%
10>Qscore>12	42,2%	10,96	12,8%
12>Qscore>15	3,6%	12,27	9,6%
15>Qscore>20	0,0%	-	-

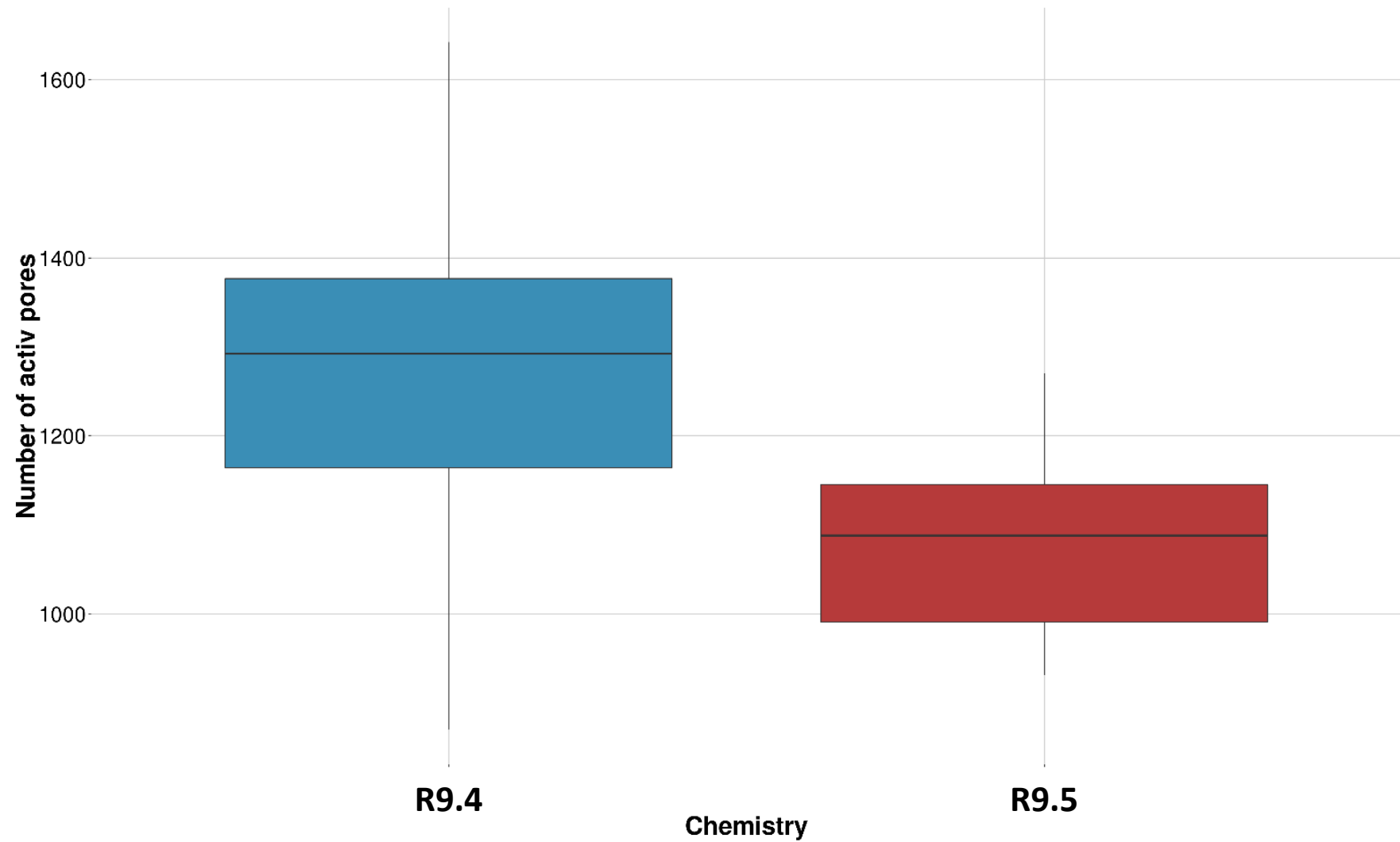


Loss of run quality for our 1D runs





Maybe related to the flowcell quality ?



Assembly : Pacbio vs Nanopore

Completeness

Example of a bacterial genome assembly :
Same quantity of data / Same informatics resources

Same contig metrics :
1 contig, 5Mb

Assemblies assesment (BUSCOv2)

Completeness based on Bacterial orthologues data base (148 genes)

Assembly sets	Complete genes	Fragmented genes	Missing genes
MinION 1D-CANU	53.4%	18.9%	27.7%
MinION 1D ² -CANU	68.9%	14.9%	16.2%
MinION 1D or 1D ² -CANU-PILON	95.3%	0	4.7%
RSII-HGAP3	95.3%	0	4.7%

Pilon : Illumina polishing

Assembly : with a more complexe genome ?

Complexity

Data	Xanthomonas campestris (5Mb)	Arabidopsis thaliana (120Mb)
Illumina	50-100 contigs	> 1000 contigs
	+ 1FC (>>>80X Nanopore)	+ 1FC (40X Nanopore)
Illumina + Nanopore	1 contig	100 contigs

Results obtained by Baptiste Mayjonade and Jérôme Gouzy

What's next at GeT ?

- ⑤ **More complexe genomes : Fishes, Fungi and Vanilla**
- ⑤ **Hybrid assembly with 10X Genomic and Illumina**
- ⑤ **Waiting for the next chemistry from Nanopore**
- ⑤ **Still testing new or updated tools (Albacore, Nanopolish) to improve the quality of the assembly with only Nanopore data**

Remerciements



**Céline Roques, Céline Vandecasteele, Claire Kuchly,
Cécile Donnadieu , Gérald Salin,
Olivier Bouchez, Alain Roulet**



Baptiste Mayjonade, Jérôme Gouzy, Fabrice Roux



Christophe Klopp



Yann Guigen, Elodie Dupin De Beyssat



Guillaume Croville, Jean-luc Guerin



Caroline Callot, Stéphane Cauet, Hélène Berges



Merci !

Expertise and results on NANOPORE, PACBIO and 10X GENOMICS technologies

Long reads : Dream or Reality

Program 2017 November 28th

9h00	Coffee
9h30-9h45	GeT Strategy Denis Milan
9h45-10h15	Implementation And Evaluation Of Oxford Nanopore MinION And GridION Sequencing Catherine Zanchetta & Maxime Manno
10h15-10h45	Minion Sequencing Provides New Insight On The Evolutionary History Of Seabird Mitochondrial Genomes Lucas Torres
10h45-11h15	Direct Whole Genome Sequencing Of Avian Poxvirus Using Nanopore MinION Guillaume Croville
11h15-11h45	INVITED SPEAKER - De Novo Assembly Of Teleost Fishes Using PacBio Sequencing Data: What Is Gained ? Ole Kristian Tørresen - Centre for Ecological and Evolutionary Synthesis, OSLO
11h45-13h15	Lunch
13h15-13h45	Implementation And Evaluation Of Chromium Technology Olivier Bouchez & Claire Kuchly
13h45-14h15	Phasing Haplotypes In Rabbit Using Long Reads Technology Julie DEMARS
14h15-14h45	High-quality De Novo Genome Assembly Of The Tomato Genome using The Latest Longs Reads Sequencing and Optical Mapping Technologies Mohamed Zouine
14h45-15h15	Two Examples Of Hard To Assemble Genomes, Even With 3rd Generation Sequences Christophe Klopp
15h15-15h45	Coffee
15h45-16h15	Not SMRT Yet Smart Sylvain Foissac
16h15-16h45	Comparison Of Methylome Profiles Between Closely Related Clones Of The Bacterial Plant Pathogen Ralstonia solanacearum Alice Guidot
16h45-17h15	Diversity Of HEV Genotype 3 Based On Full-length Sequences Florence Nicot

INRA Get-PlaGe
Salle de Conférence Marc Ridet
24 Chemin de Borde Rouge, 31326 Castanet-Tolosan

TOULOUSE
EUROPEAN CITY OF SCIENCE
ESOF 2018

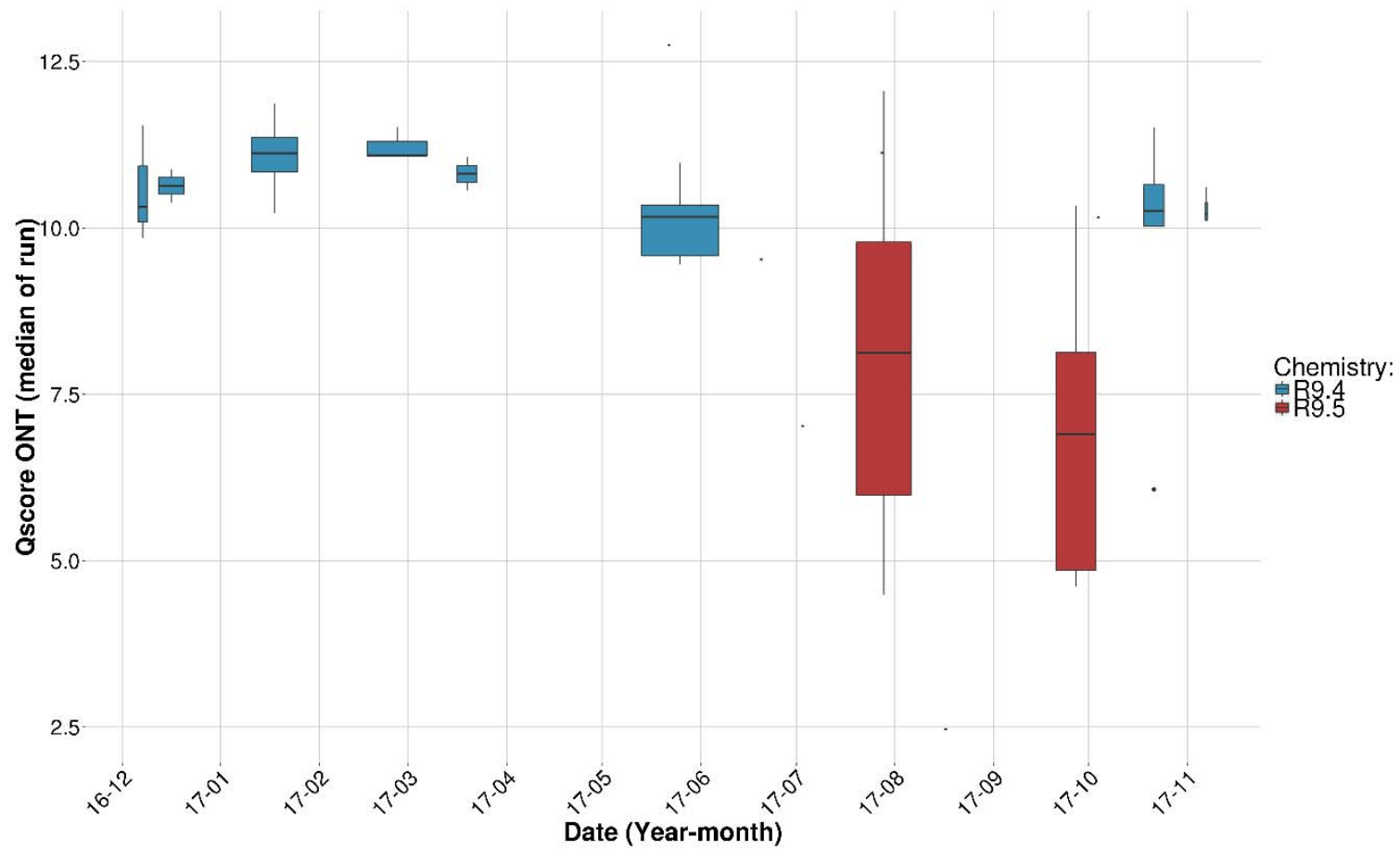
INRA
SCIENCE & IMPACT

Genotoul GeT

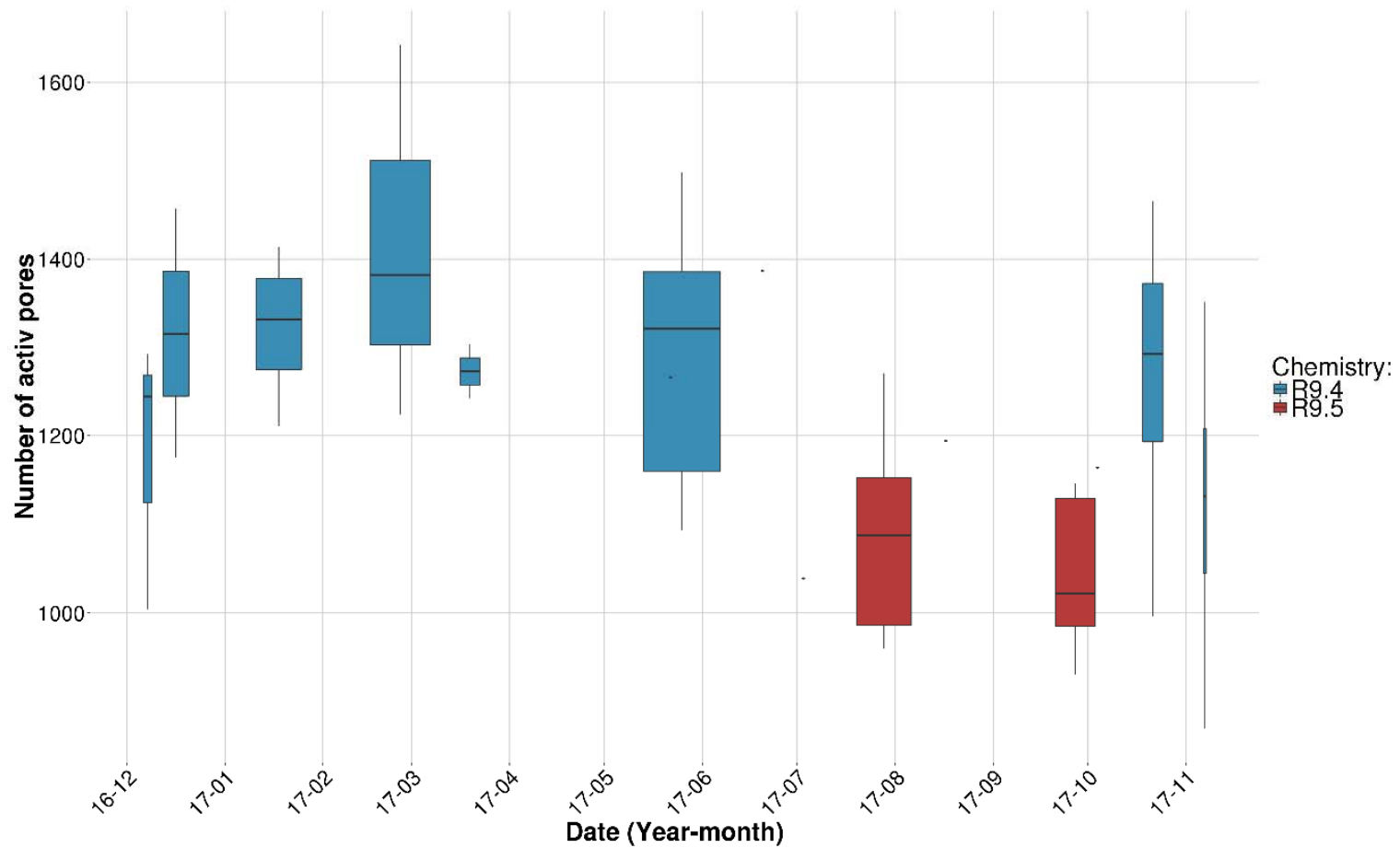
The biennial European event, EuroScience Open Forum (ESOF) will take place in Toulouse, "European City of Science" from 9 to 14 July 2018. "Sharing science: towards new horizons" will be the motto of this ESOF 2018 edition, a 6-in-1 event including various sections: "Science", "Science policy", "Science to Business", "Careers" and "Media & Science Communication" and a "Science in the City" programme dedicated to the general public. A series of themes covering all fields of science and their relations with society are covered by this multidisciplinary event through conferences, exhibitions and satellite events.



Qscore of 1D runs per month



Related to the flowcell quality ?

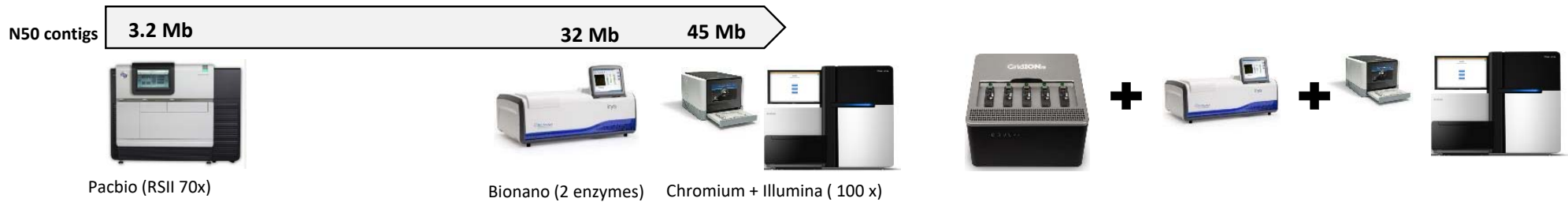


To conclude !!!

- § To go further and improve genomes assembly : scaffolding technologies (10X Genomics, BioNano...)

Tomato genome 800Mb (end 2016)

Fish genomes 1-2Gb (end 2017), Vanilla 4Gb (2018)



- § ONT has opened the door to a wide range of applications

- § Nanopore technology : Long DNA fragments = long reads

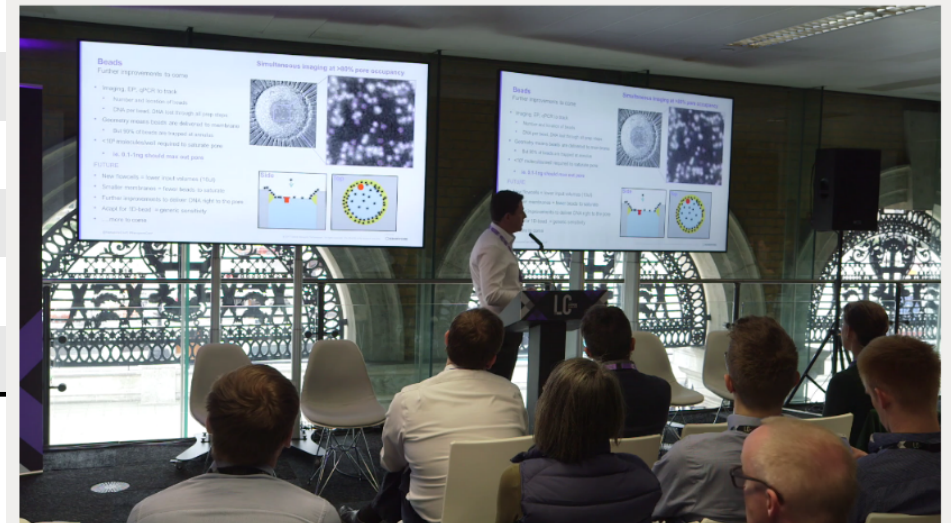


Pacbio vs Nanopore

Comparison

	Nanopore	Pacbio
		RSII SEQUEL (expected)
Sensitivity (DNA quality)	Sensitive	Very sensitive
Lib prep	2 days	4 days
Yield	7.5Gb/FC	
Bioinformatic analysis	3-6 days	
Methylations		
Assembly quality		
Price		+1
DNA quantity	1.8µg/Gb (5.4 for 1D ²)	

London Calling 2017





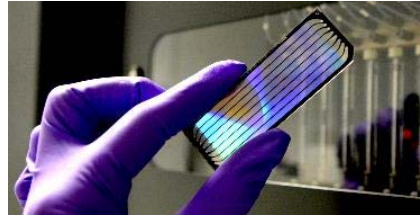
Genome assembly using Nanopore technology

Assemblies improvement (small genomes)

Complexity →

Data	<i>Xanthomonas campestris</i> (5Mb)	<i>Arabidopsis thaliana</i> (120Mb)	<i>Ganoderma boninense</i> (50Mb)
Illumina	50-100 contigs	> 1000 contigs	> 10 000 contigs
	+ 1FC (>>>80X Nanopore)	+ 1FC (80X Nanopore)	+ 1FC (100X Nanopore)
Illumina + Nanopore	1 contig	140 contigs	525 contigs

Short reads



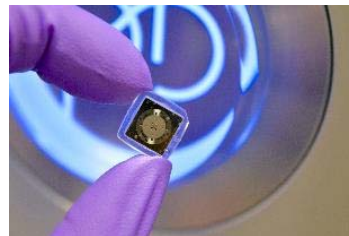
Illumina : MiSeq + HiSeq 3000

Synthetic long reads



10x genomics : Chromium

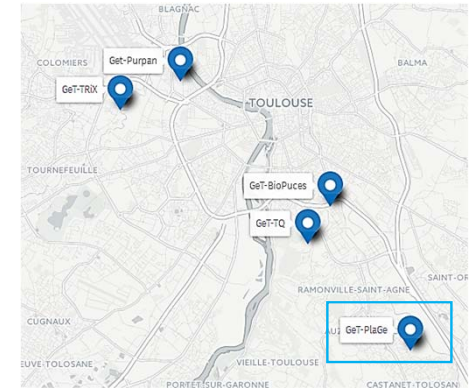
Long reads



Pacbio : RSII

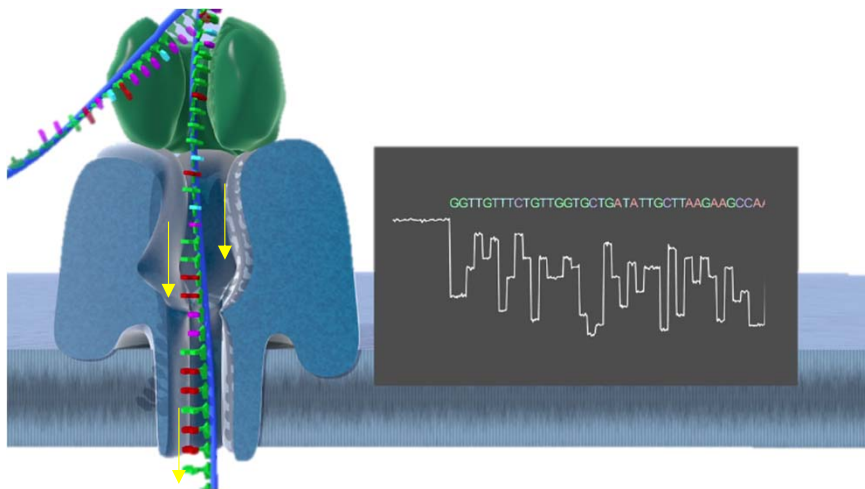
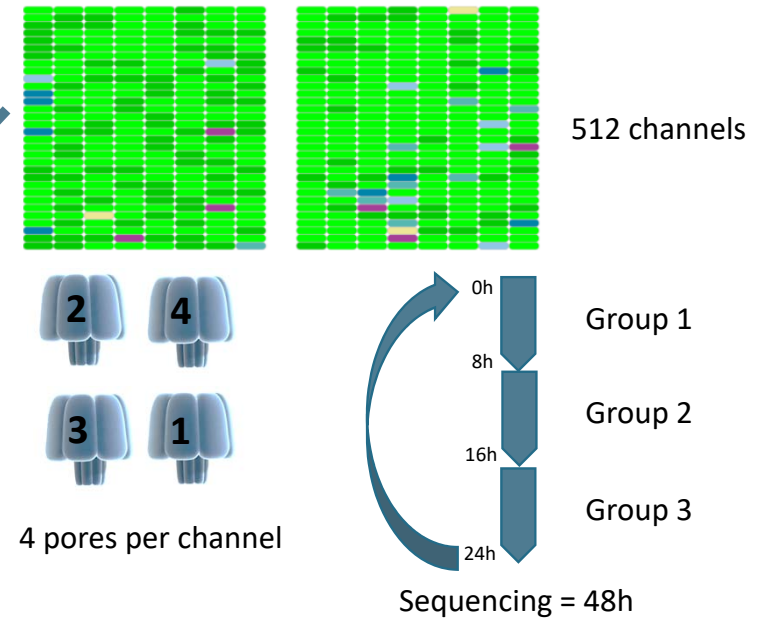


ONT : MinION + GridION



Nanopore technology

How it works



- A **protein** is set in an electrically resistant polymer membrane
- An **ionic current** is passed through the nanopore
- The **event / base** creates a characteristic disruption in current
- Identification of G, A, T and C bases

What can we expect ?

The yield might be very different from a flowcell to another

Good quality DNAs + 15-20kb shearing



Similar number of molecules loaded

Gb/FC

