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

Journée Long Reads GeT

28 Novembre 2017

Catherine Zanchetta & Maxime Manno

http://get.genotoul.fr
get@genotoul.fr
@get_genotoul

Genotoul GeT

## Wet Lab

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

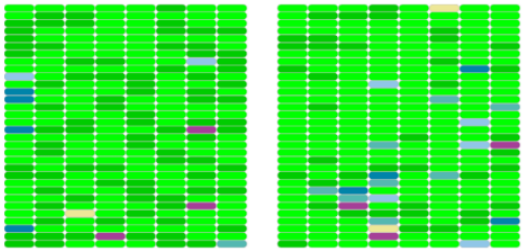
## Bioinformatics

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

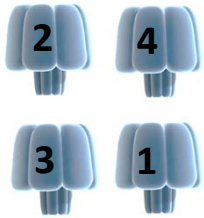
Genotoul GeT

# Nanopore technology

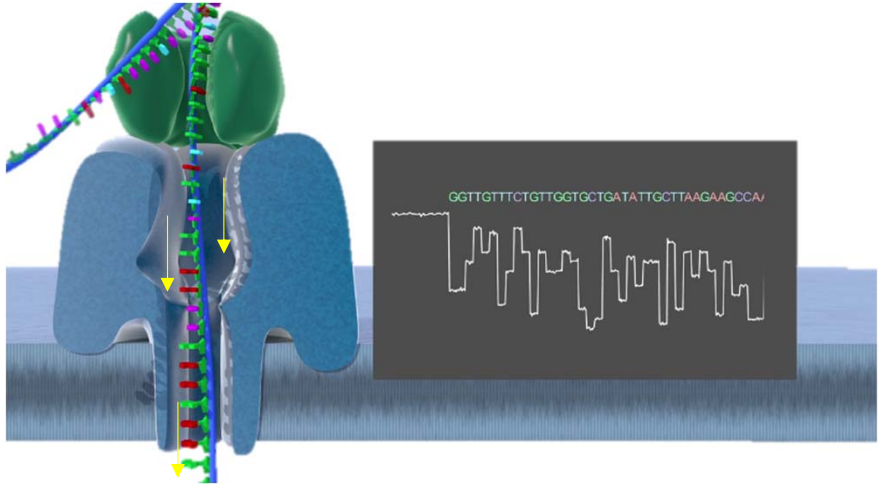
*How it works*

512 channels




4 pores per channel

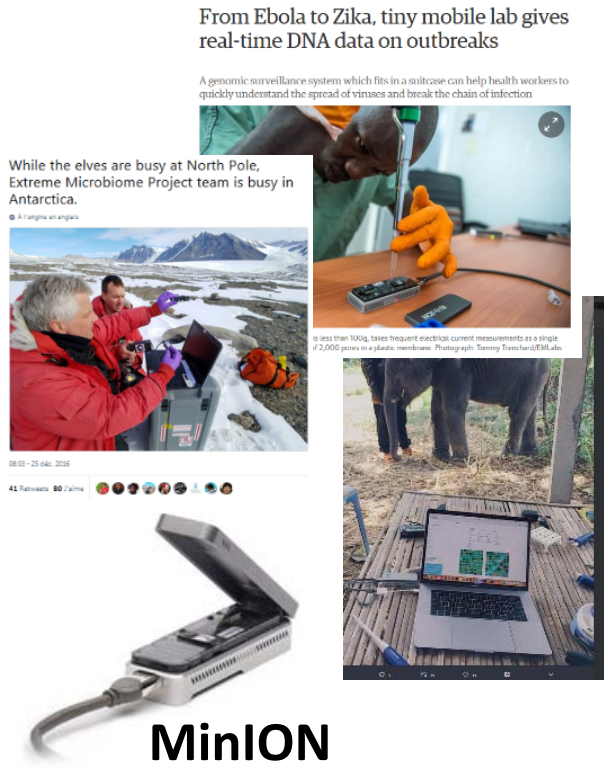



- A **protein** is set in an electrically resistant polymer membrane
- An **ionic current** is passed through the nanopore
- The **event / base** creates a characteristic disruption in current
- Identification of G, A, T and C bases

Genotoul GeT

# Nanopore technology

*A scalable technology*




**MinION**

~ 7Gb / 48h

**GridION**

~ 35Gb / 48h

**PromethION**

~ 2Tb / 48h ???

2016 → 2017 → 2018

Genotoul GeT

# Nanopore technology

*MinION and GridION*

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



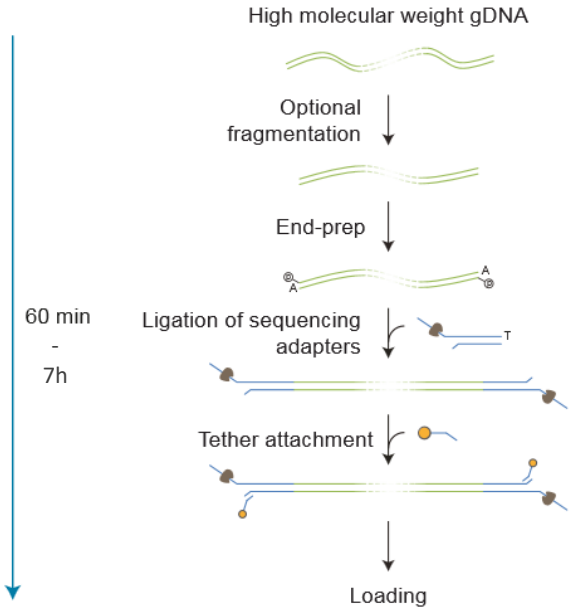
| | | |
|---|---|---|
| **Number of FC** | 1 | 5 |
| **Live basecalling** | We don't use it | We use it ! |
| **Basecaller version** | Last version available | Delay for the incorporation of the last version into dogfish |
| **Price per FC** | 675€ | 300€ |

Genotoul GeT

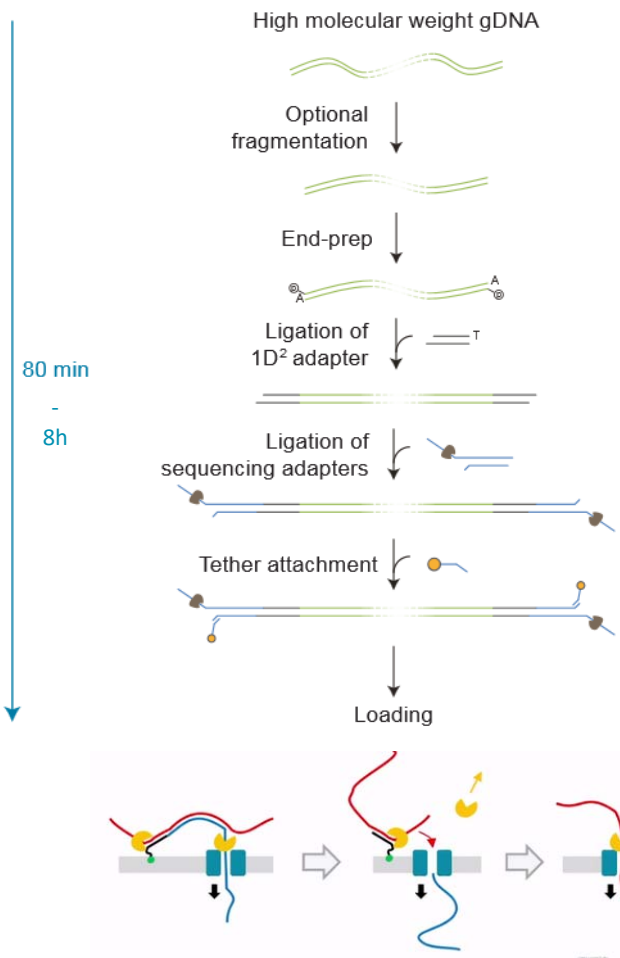
# Library preparation

*1D and 1D² kits*

## 1D




## 1D²



# Library preparation

*Importance of the quality controls*

- Have a **good quality DNA**

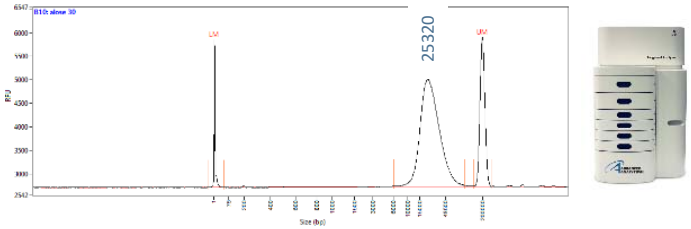
Nanodrop

A260/280 = 1.8-2.2

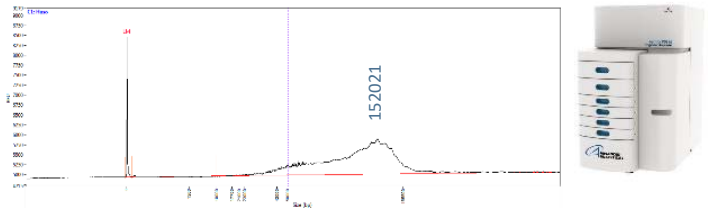
A260/230 = 1.8-2.2

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

Qubit




Fragment analyzer
DNA < 40 kb




FEMTO
DNA > 40 kb

Genotoul GeT

# Library preparation

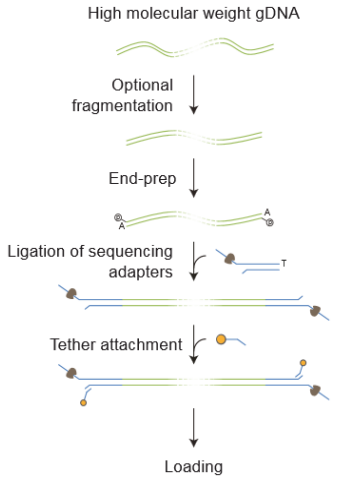
*Suitable library prep*

- 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)



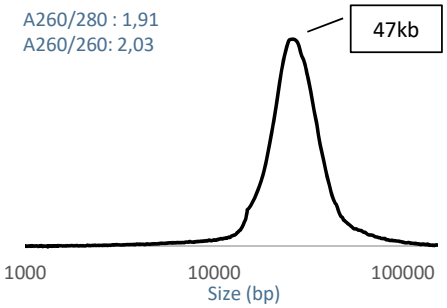

1.8µg / Gb of data

# What can we expect ?

*The read distribution depends on the DNA quality*

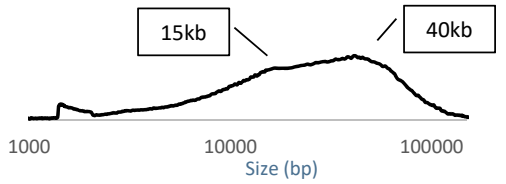
## Non degraded DNA

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



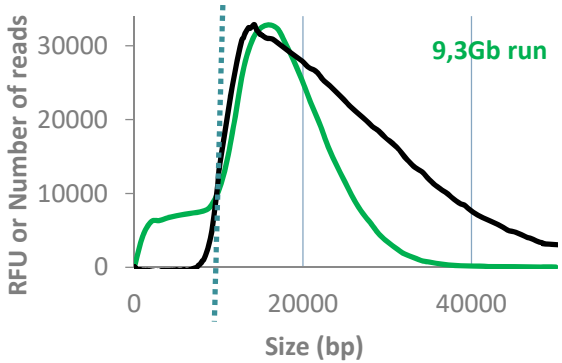

## Degraded DNA

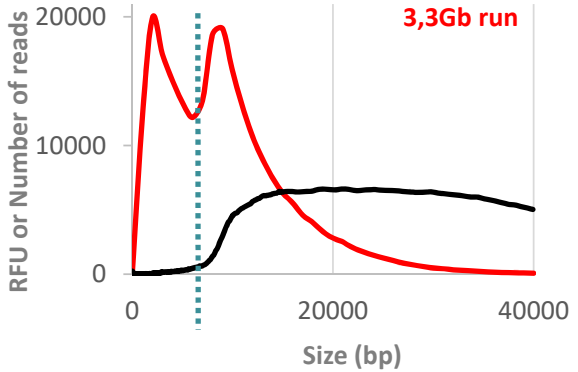
A260/280 : 1,94
A260/260: 1,97




20kb shearing

>10kb selection




Library  Reads  Blue Pippin Cutoff

Genotoul GeT

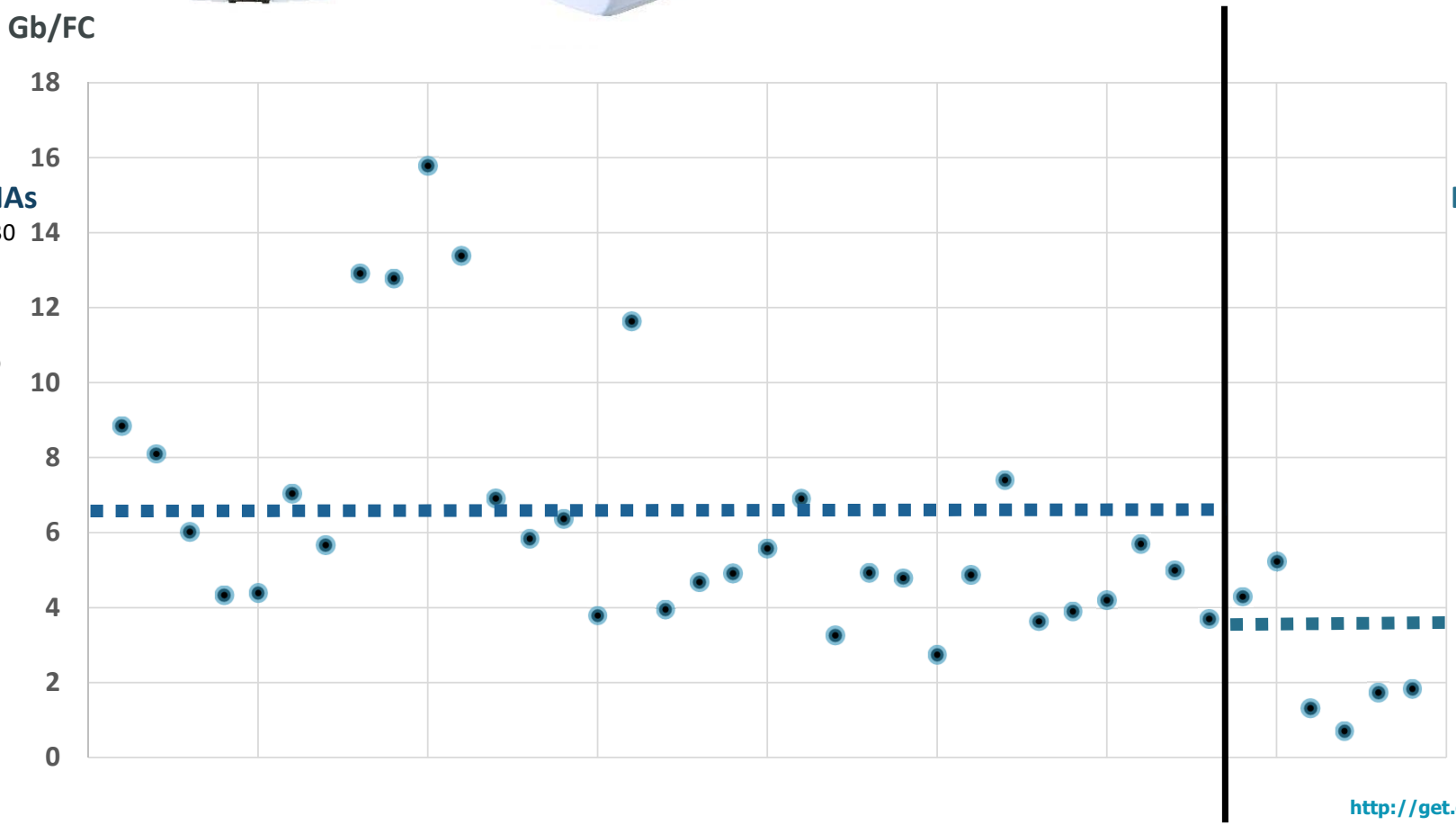
# What can we expect ?

*The yield might be very different from a flowcell to another*

15-30 kb

Only if necessary

pM Similar number of molecules loaded




**Good quality DNAs**
A260/280 ; A260/230
1.8-2.2

Mean yield : 6.5 Gb
~ 70 % reads >Q10

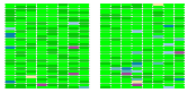
**Bad quality DNAs**
A260/280 ; A260/230
<>1.8-2.2

Mean yield : 2.5 Gb
~ 40 % reads >Q10

Genotoul GeT

# What can we expect ?

*15kb – 20kb DNA : prefered size*

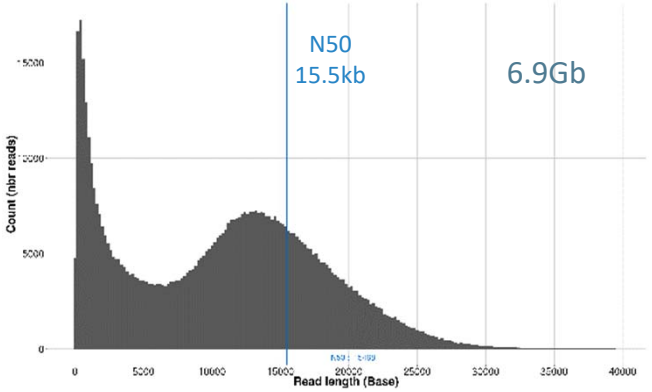


Similar number of active pores

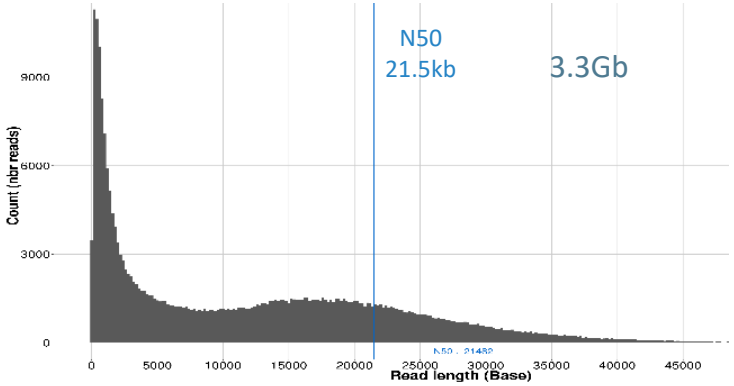
pM Similar number of molecules loaded

DNA 1

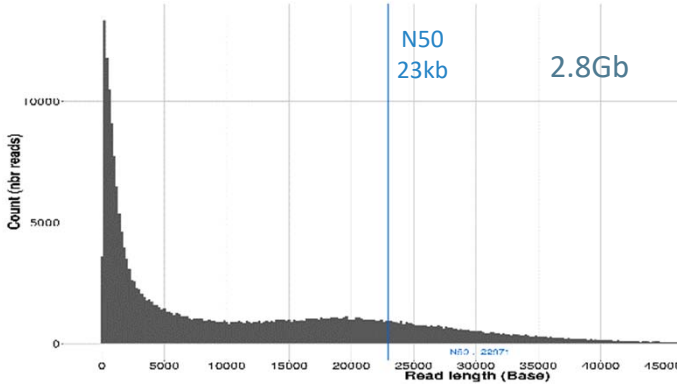
## 20kb shearing




## 30kb shearing



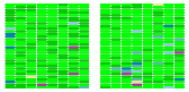

## 40kb shearing



Genotoul GeT

# What can we expect ?

*15kb – 20kb DNA : prefered size*

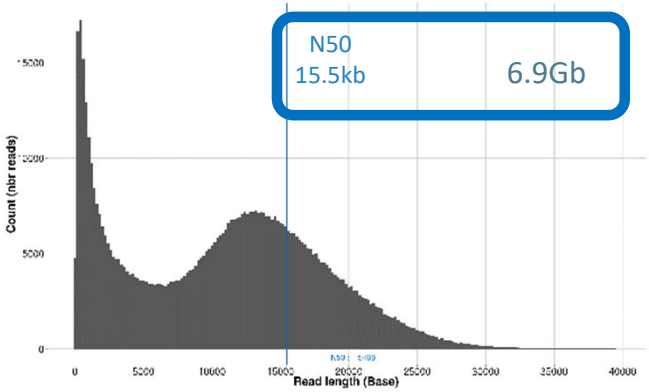


Similar number of active pores

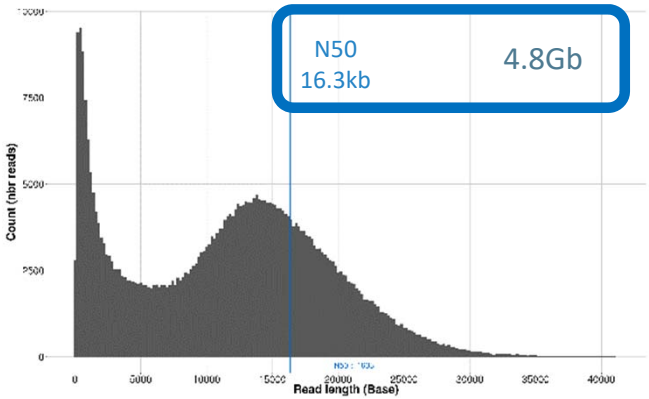
pM Similar number of molecules loaded

## 20kb shearing

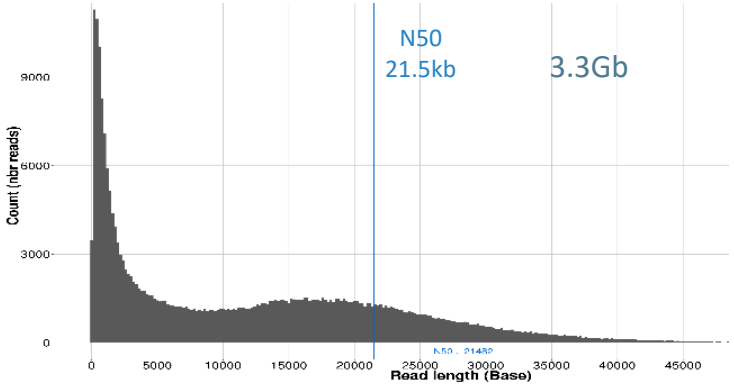
DNA 1

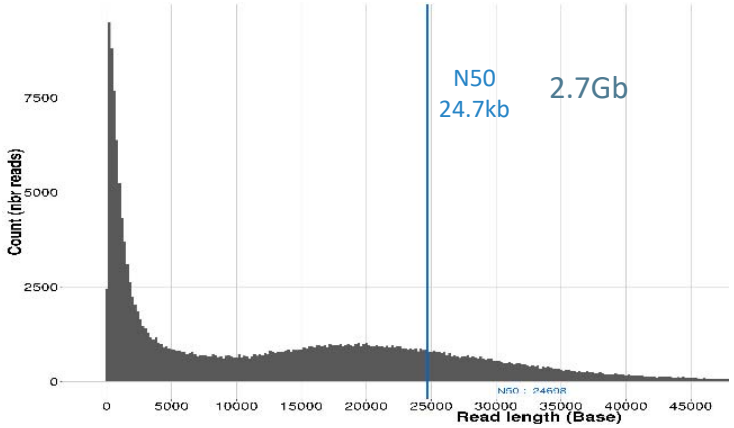



DNA 2

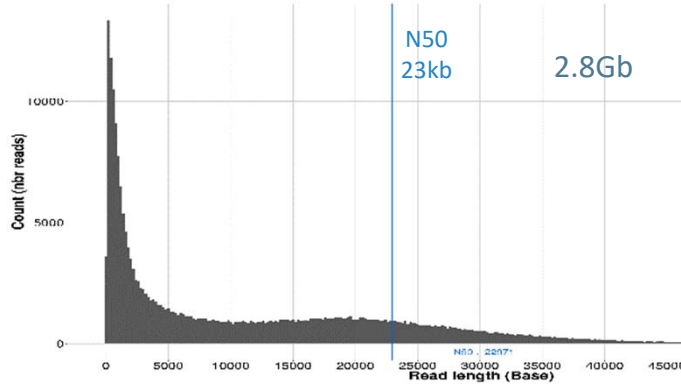


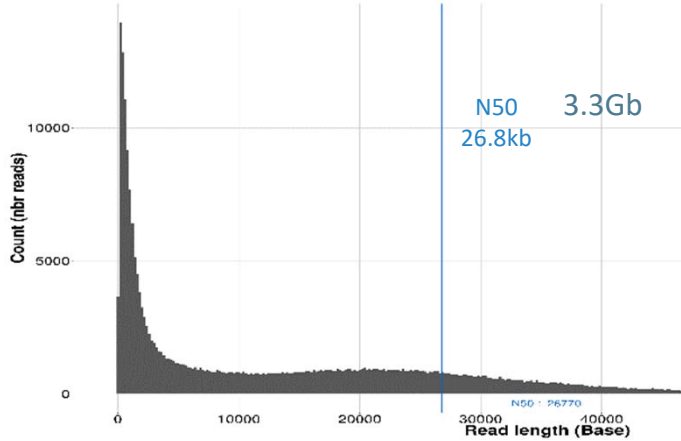

## 30kb shearing




## 40kb shearing

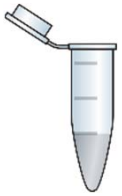


Genotoul GeT

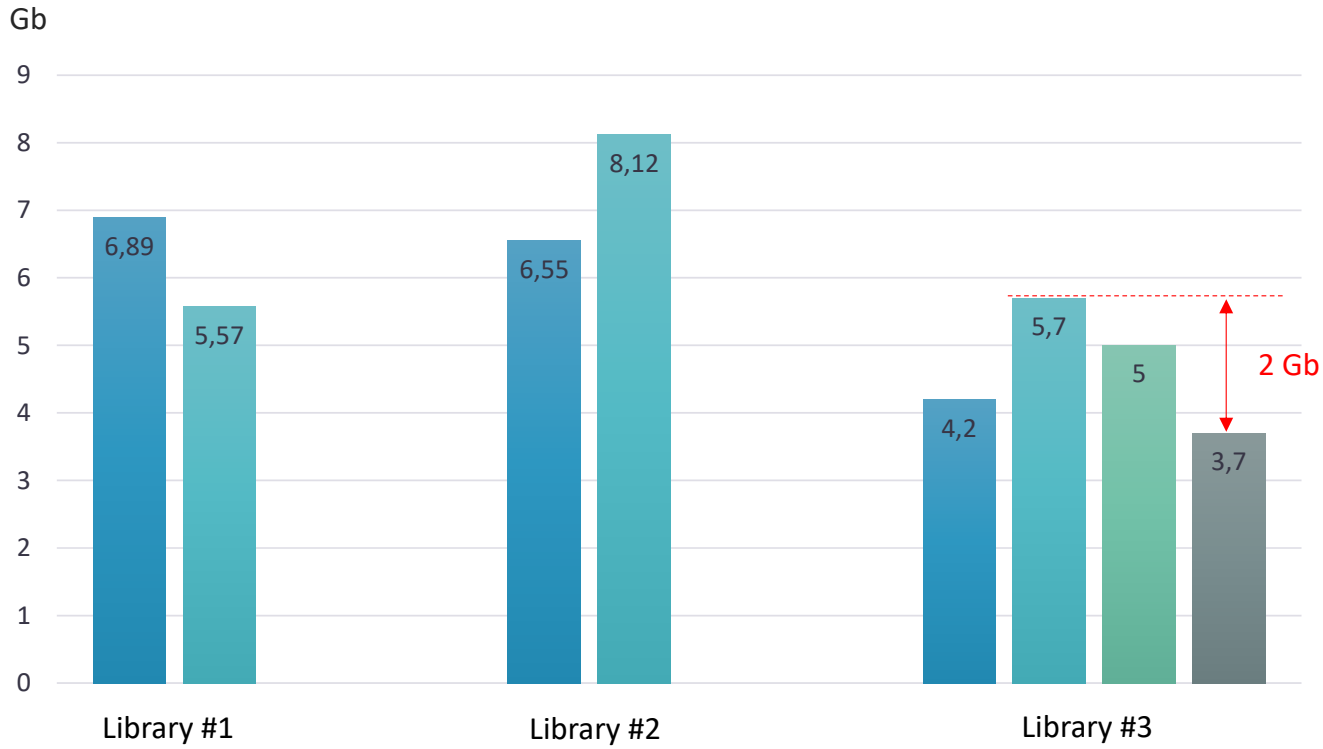
# What can we expect ?

*Yield variability between flowcells*



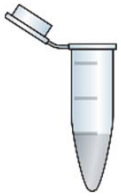
1 library

Many FC (R9.4)



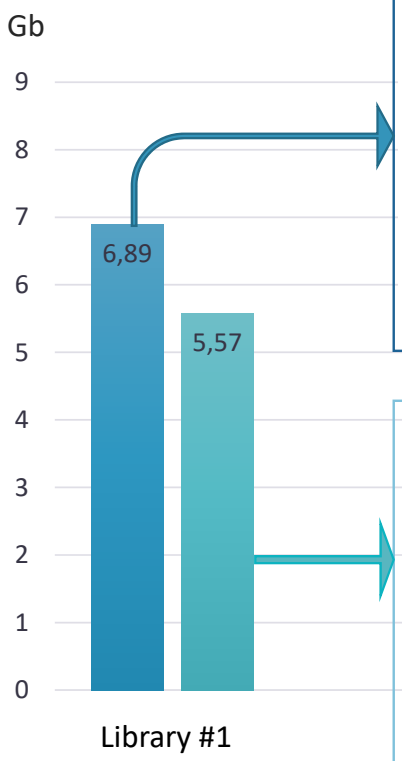
# What can we expect ?

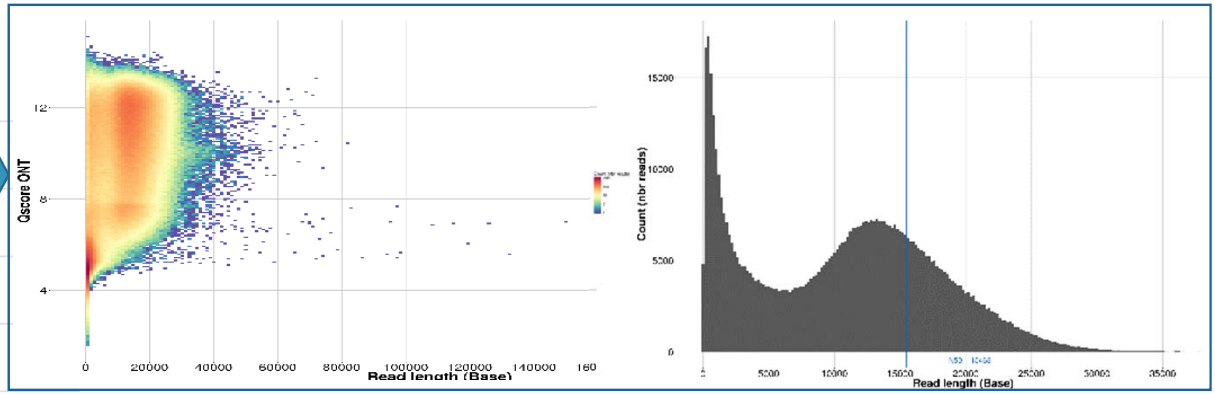
*Yield variability between flowcells*

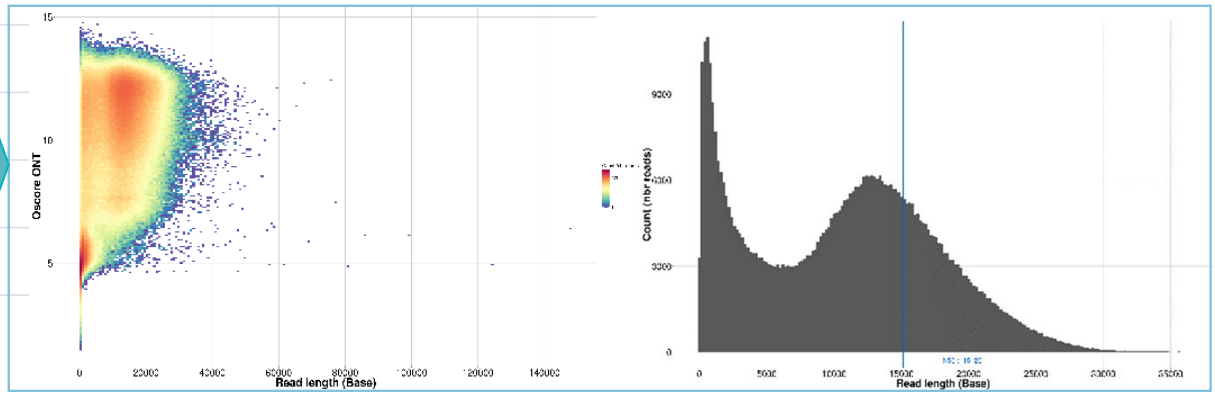


1 library

Many FC (R9.4)



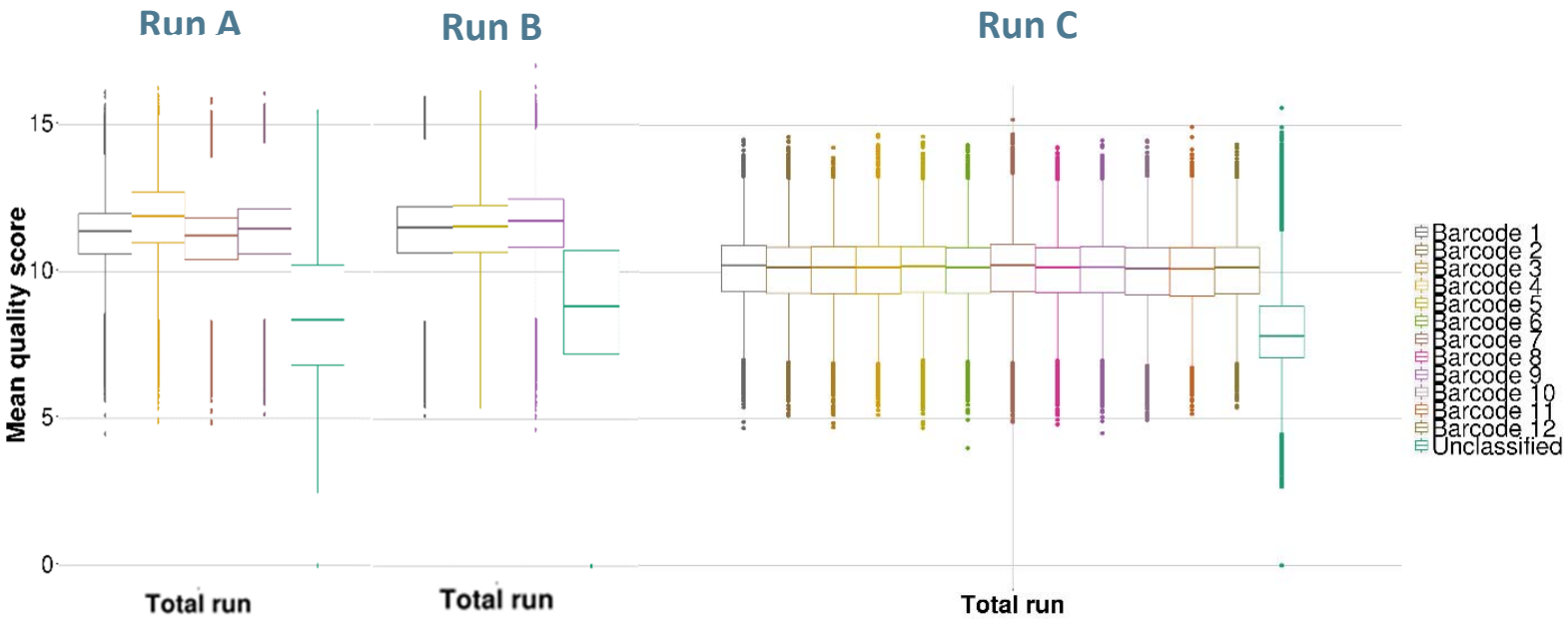



Genotoul GeT

# What can we expect ?

*Multisample run to lower the cost*




3 and 4plex - 8kb shearing

12plex – 20kb shearing
40 000 – 76 000 per sample

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

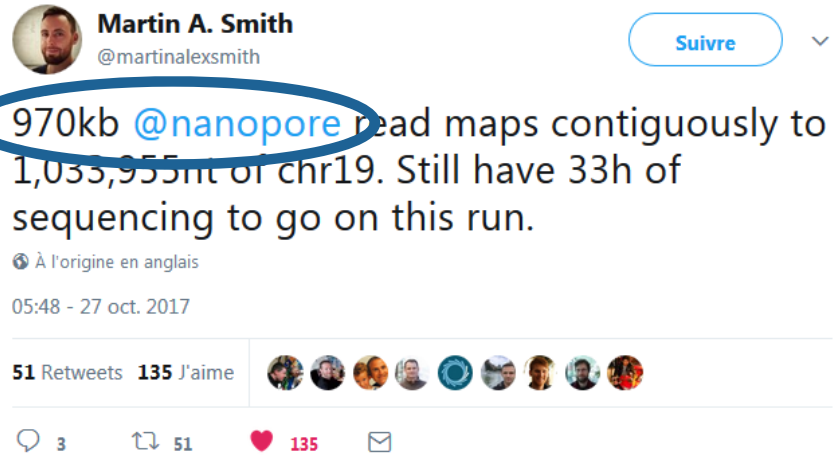
A et B : R9.4 FC / 3Gb runs / Minknow 1.5 / Albacore 1 ; C : R9.4 FC / 13.4Gb run / Minknow 1.7.3 / Albacore 1.1.0

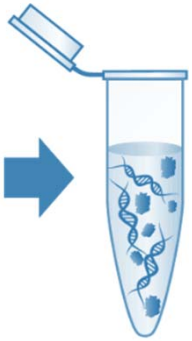
Genotoul GeT

# What can we expect ?

*Longest read : 1Mb read*

**Rapid protocol**



**Extraction : key step**

Genotoul GeT

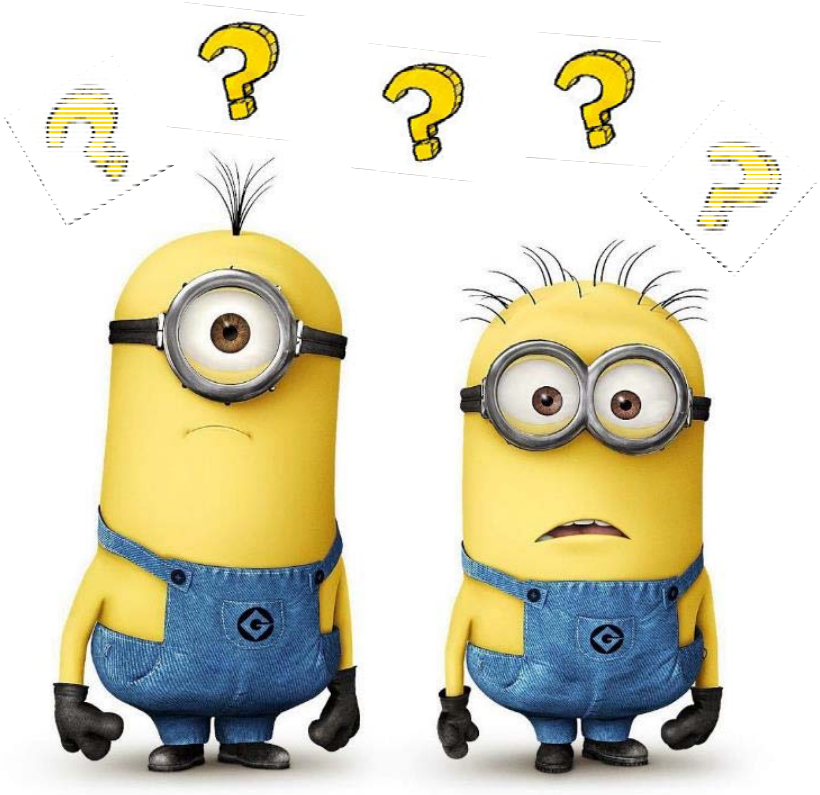
- **Wet Lab**
  - Nanopore technology
  - Library preparation
  - What can we expect ?

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

Qscore ONT

Dogfish

ONT data



Minknow

Albacore

Porechop

Polishing

MinION

Fast5 file

Easy ???

What is it ? How to do it ?

Genotoul GeT

# Thanks to :

Strong collaboration with the Genotoul Bio-informatics Plateform.

- IT resources
- Great help

Genotoul Bioinfo

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

Genotoul GeT

# IT solution : Sequencing

Oxford NANOPORE Technologies


**Location :**
**MinION Computer/ GridION**
**Genotoul Server**

**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

# 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)

Fastq

Porechop

https://github.com/rrwick/Porechop

Filtlong

**FILTERING – Homemade tool (or Filtlong)**

→ Output : filtered reads

Size > 3 kb

Qscore > 10

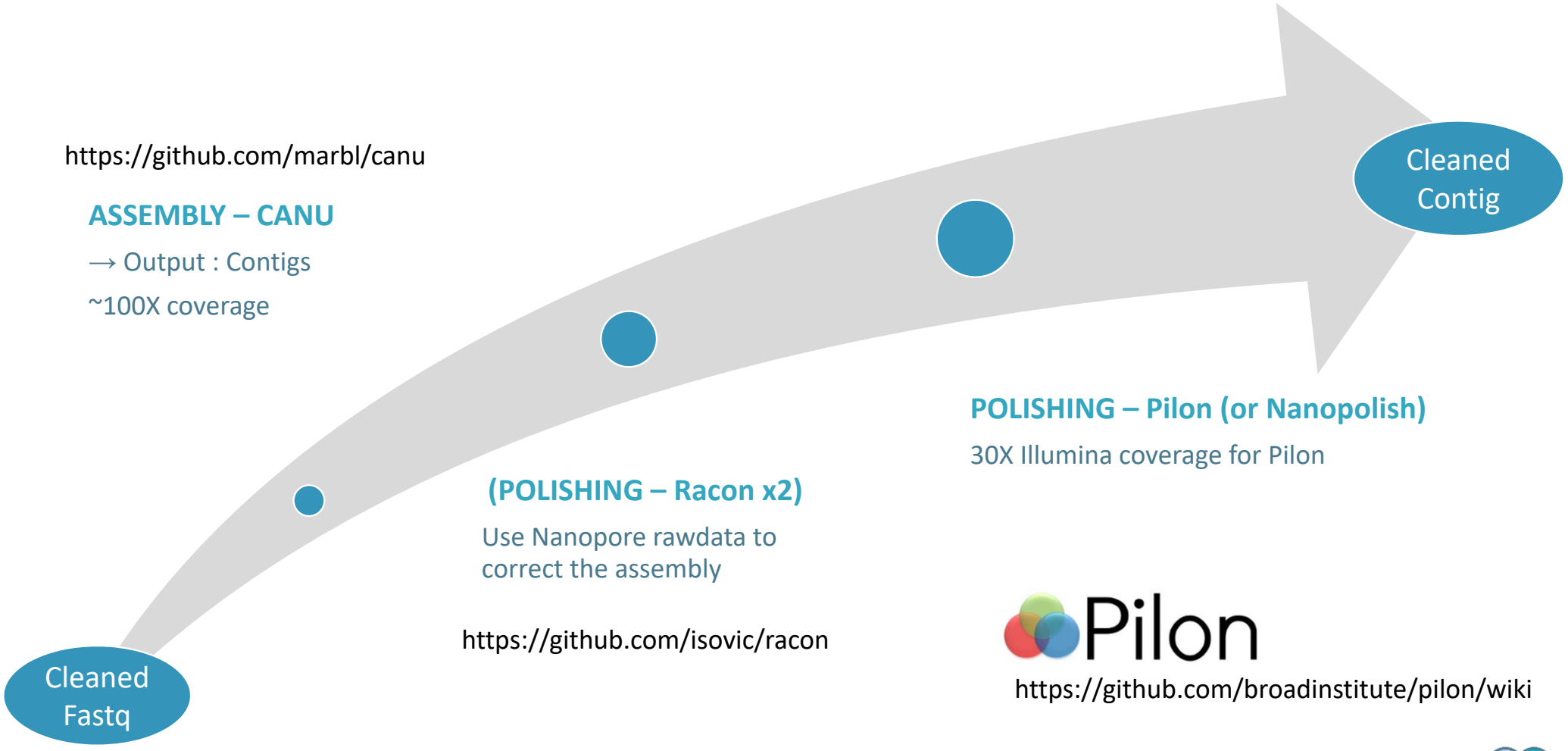
https://github.com/rrwick/Filtlong

Cleaned FASTQ

Genotoul GeT

# Bio-informatic solution : Assembly and polishing

*Bacterial genome (5Mb)*



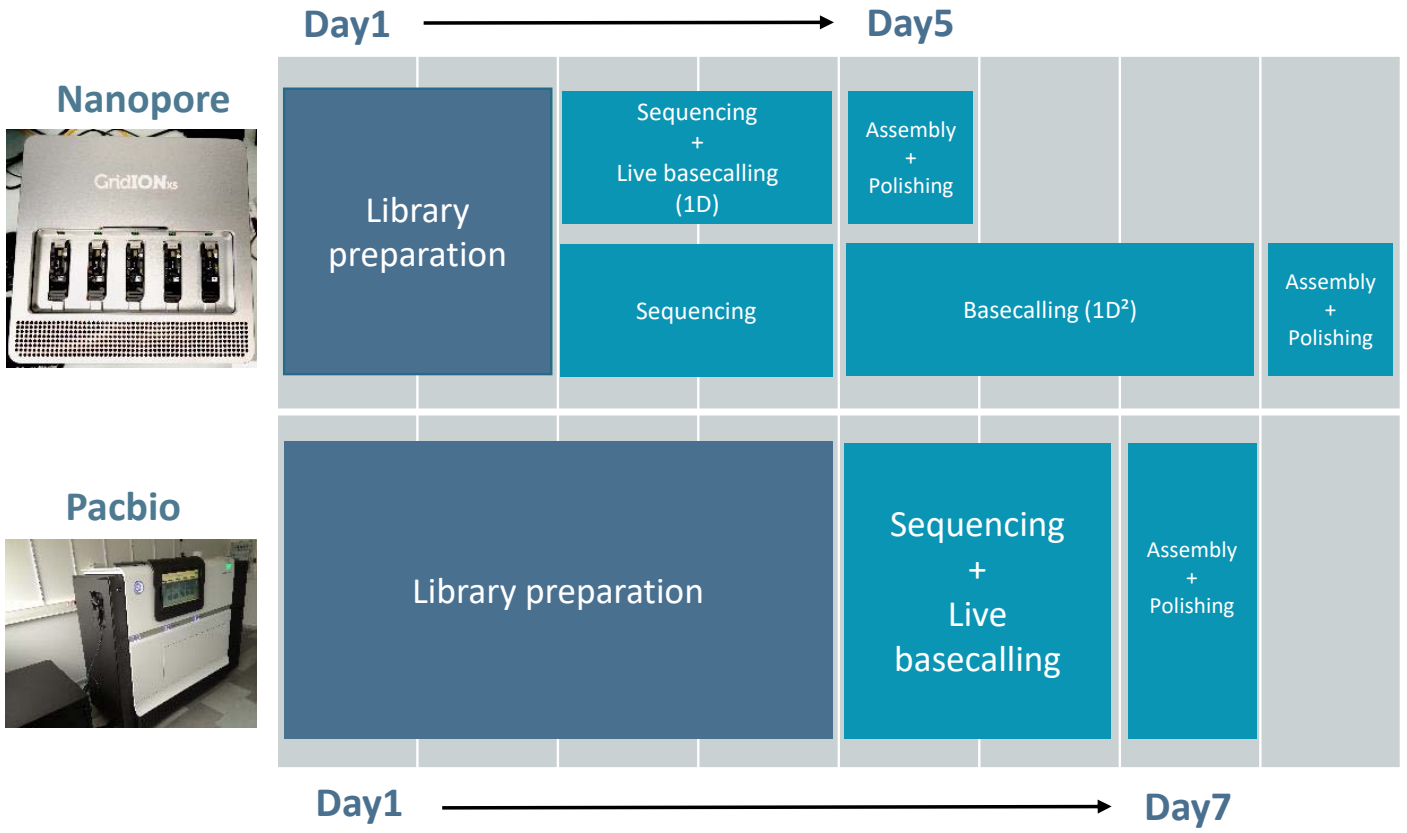
- IT and Bio-informatic solution
  - Sequencing and raw data (MinION, GridION)
  - Trimming and filtering
  - Assembling and polishing
- **Some results about bacterial genome sequencing**
  - **Deadlines and computing time**
  - **Raw data : Error rate and Qscore**
  - **Assembly : Completeness**

Genotoul GeT

# 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*

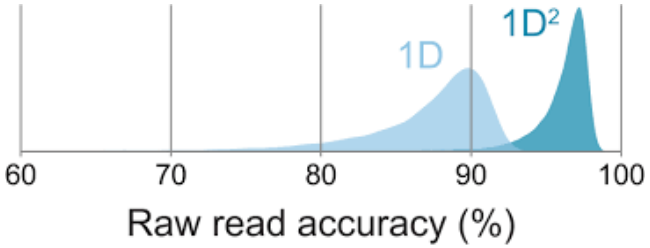
## Raw read error rate

**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 1D²** (ONT filtered reads) | 8.8 % |

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

Oxford Nanopore announcement




**Qscore ? (Quality score for ONT data)**

Genotoul GeT

# 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

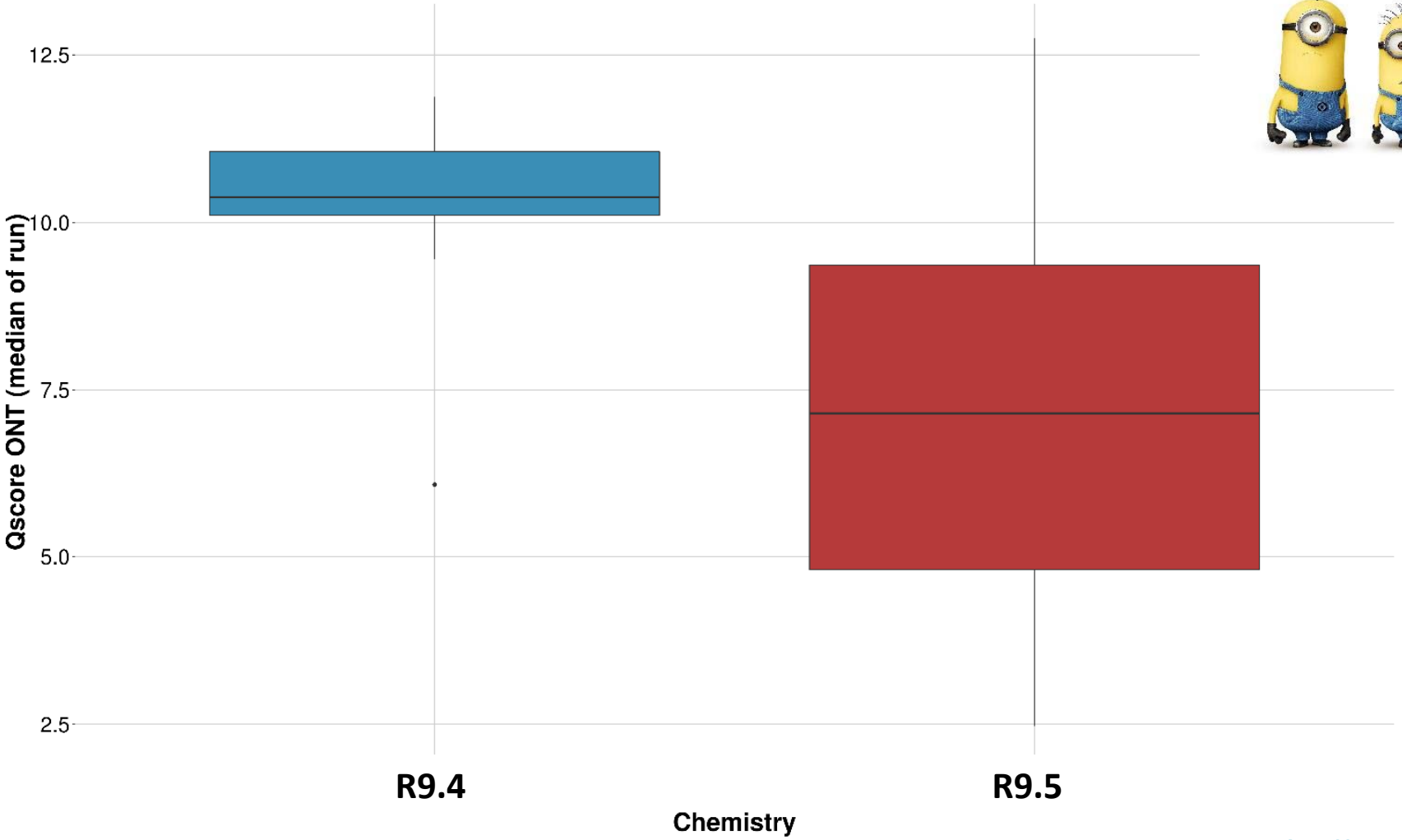
| Subset | %reads | Mean Qscore | Error rate |
|---|---|---|---|
| 0>Qscore>5 | 0,1% | 4,75 | 45,7% |
| 5>Qscore>8 | 11,6% | 7,05 | 34,4% |
| 8>Qscore>10 | 17,7% | 9,13 | 23,7% |
| 10>Qscore>12 | 42,3% | 11,14 | 15,1% |
| 12>Qscore>15 | 28,3% | 12,44 | 10,7% |
| 15>Qscore>20 | 0,0% | - | - |

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% | - | - |

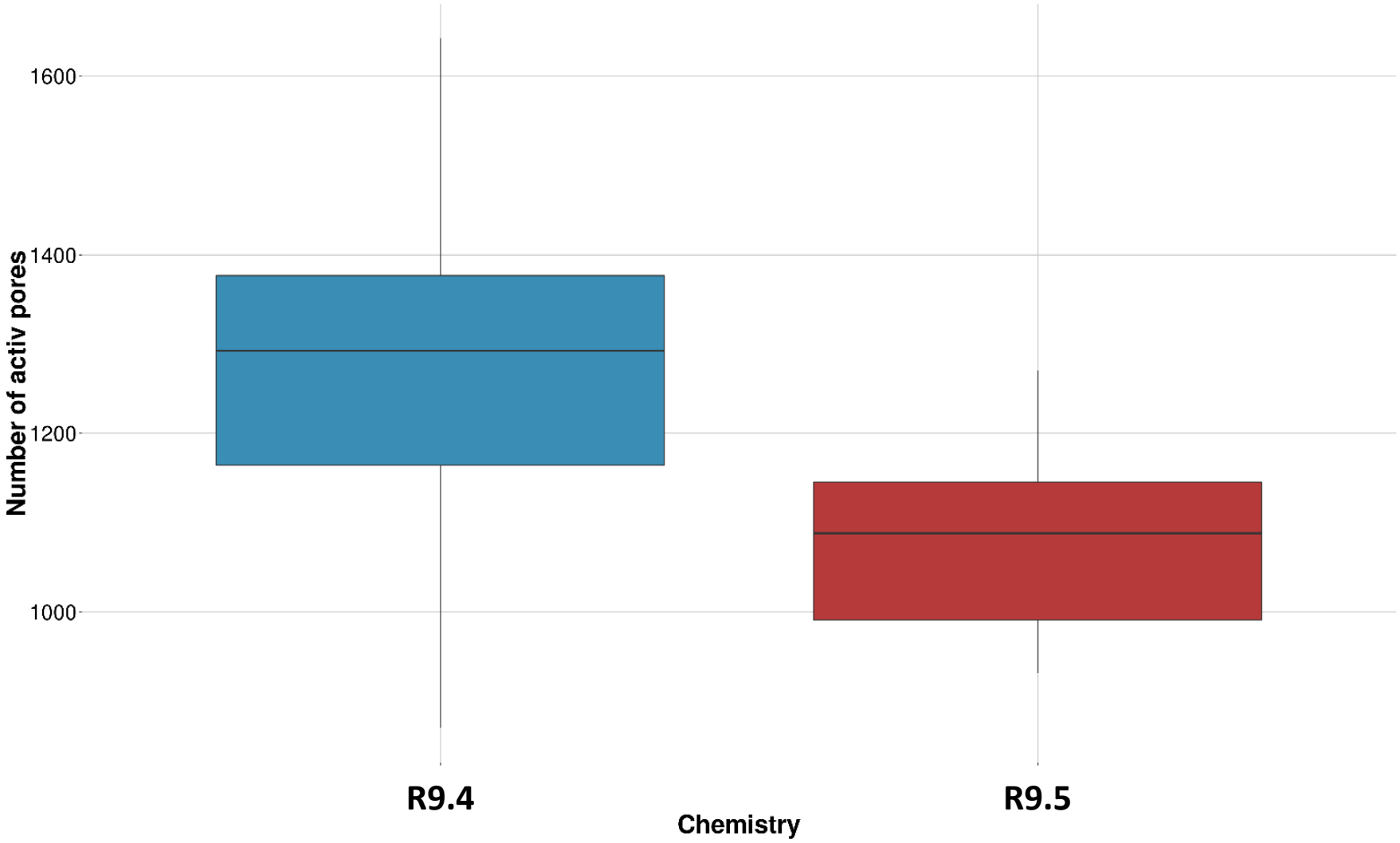
Genotoul GeT

# Loss of run quality for our 1D runs



Genotoul GeT

# Maybe related to the flowcell quality ?



Genotoul GeT

# 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)

Completness 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 1D²-CANU | 68.9% | 14.9% | 16.2% |
| MinION 1D or 1D²-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

Genotoul GeT

# What's next at GeT ?

- More complexe genomes : Fishes, Fungi and Vanilla
- Hybrid assembly with 10X Genomic and Illumina
- Waiting for the next chemistry from Nanopore
- 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

Genotoul GeT

# Merci !