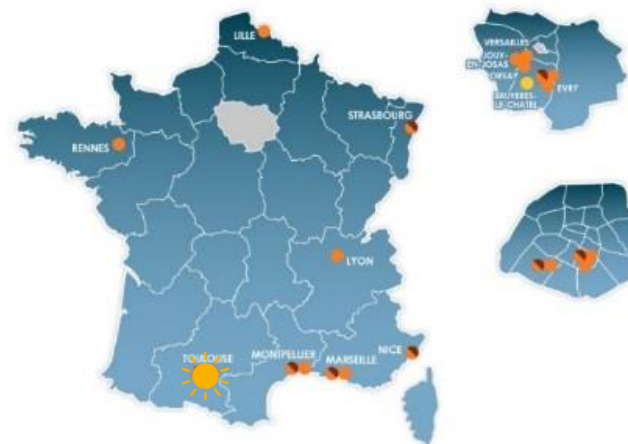


From DNA preparation to the assembly, trying to get the best of the MinION

Catherine Zanchetta & Maxime Manno

France Génomique Bordeaux 22-23 Juin 2017

- ⑤ **Genomics and Transcriptomics (GeT) Platform of Genotoul hosted by**



- ⑤ **A strong partnership with a bioinformatics core facility**



- ⑤ **A node of the National Distributed Infrastructure « France Génomique »**



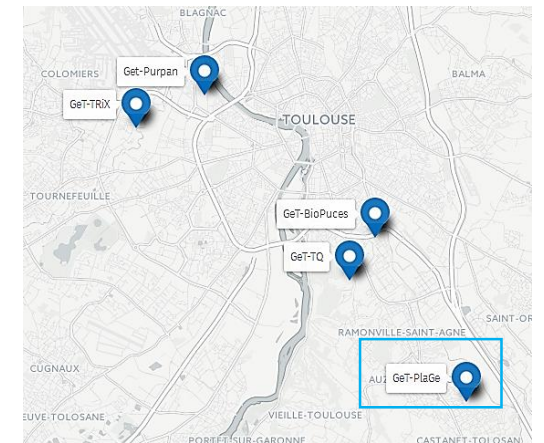
- ⑤ **Quality certifications ISO9001 & NFX 50 900, Propel**



Team and expertise

③ 35 people on 5 sites

- Experts in Agronomy, Environment, Microbiology, Health
- Competence in biology, bioinformatics, biostatistics



⑤ **More than 140 laboratories in 2016** (INRA, CNRS, INSA, INSERM, CHU, CIRAD ...)

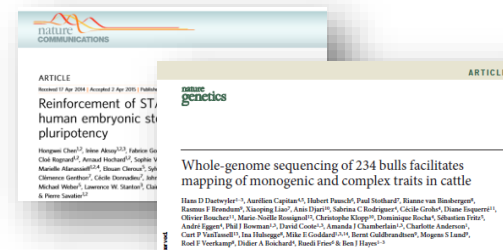
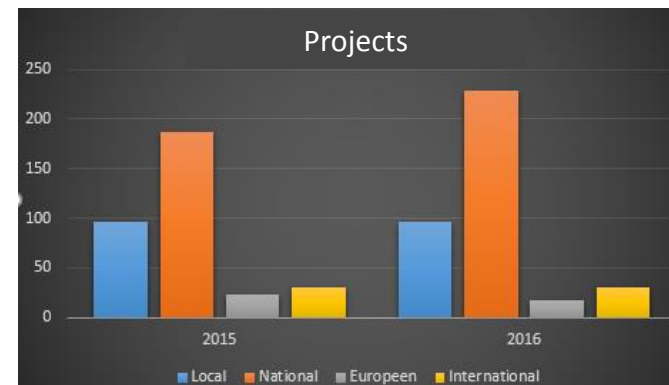
- More than 240 research teams
- More than 360 projects
- >2M€ of activity

⑤ **10 R&D projects**

- Chromium, MinION, ChIPseq, methylation, HiC...

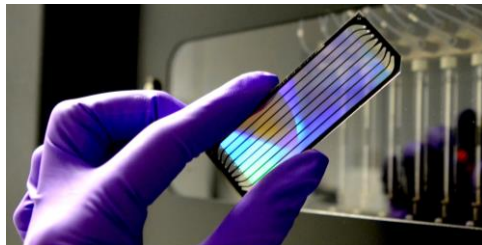
⑤ **19 Research projects (ANR, INCA, H2020 ...)**

⑤ **91 Publications as co-author since 2010**

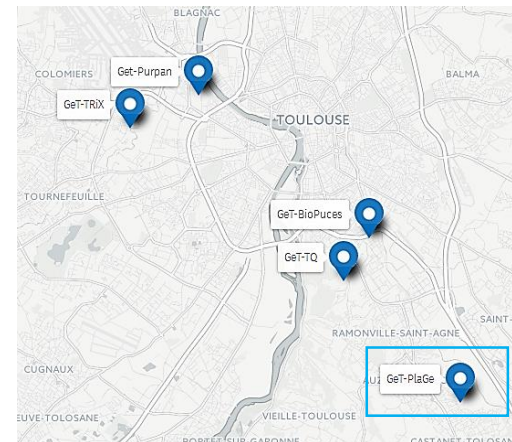




Short reads



Illumina : MiSeq + HiSeq 3000



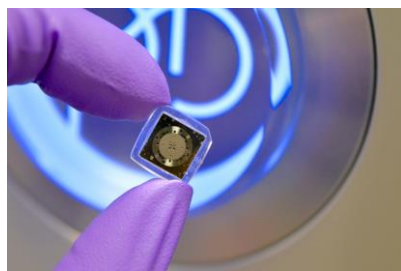
Synthetic long reads



10x genomics : Chromium

Pauline Heuillard
Whole genome sequencing
de 10X Genomics : les
premiers tests

Long reads



Pacbio : RSII

Céline Vandecasteele
Direct detection of DNA
methylation during SMRT
sequencing with PacBio
RSII



ONT : MinION

Baptiste Mayjonade
Exemples d'une bactérie
phytopathogène et
d'*Arabidopsis thaliana*



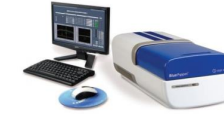
ONT offer

Library preparation kits for DNA

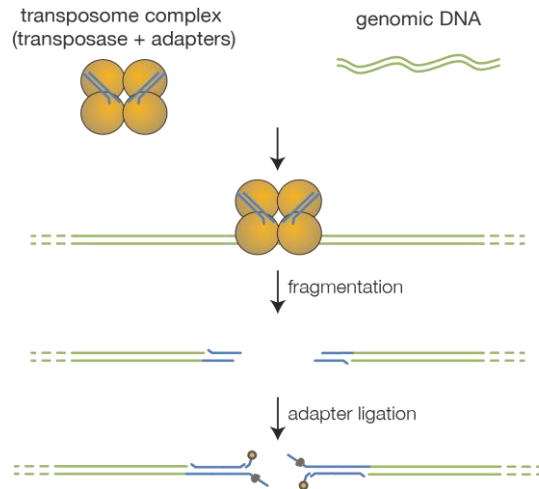
Shearing

Size selection

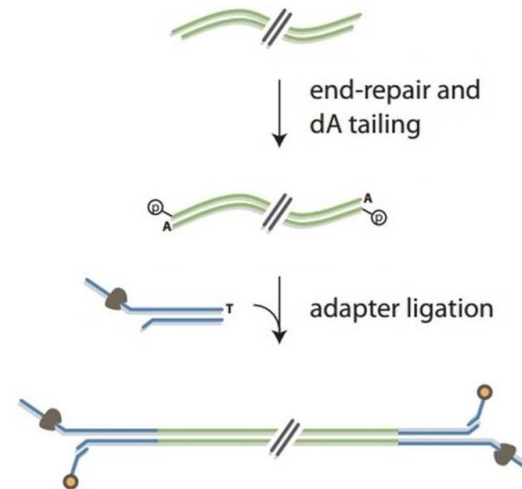
Optional



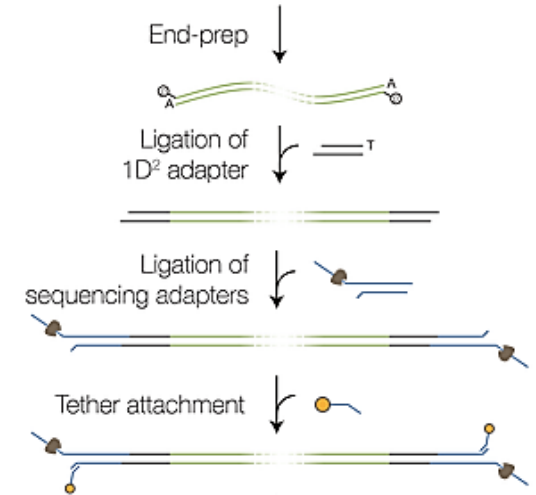
Rapid kit



1D kit



1D² kit



+

Fast lib prep : 10 minutes
Very long reads

1-2 days

High yield (2-15 Gb / FC)

1-2 days

High yield (2-15 Gb / FC)

-

Low yield (200 – 500 Mb)

Important amount of DNA required

% of sequences in 1D² ?

Important amount of DNA required



Library preparation

Good practices

§ Shearing



gtube

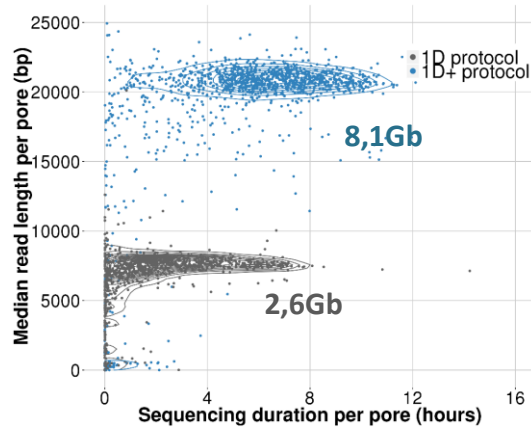


Megaruptor

§ Size selection



Blue Pippin

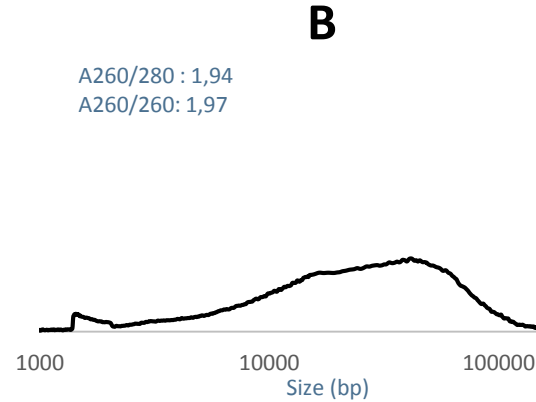
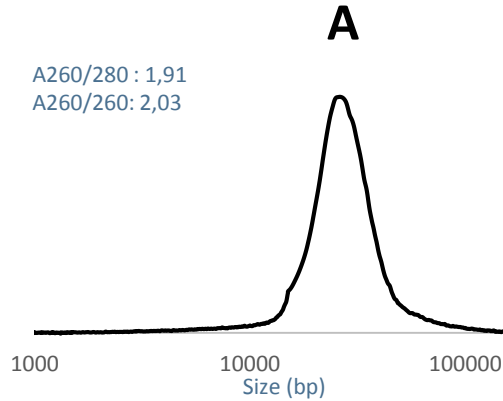


§ Long fragments : suitable library preparation

Depends on the biological question and on the DNA

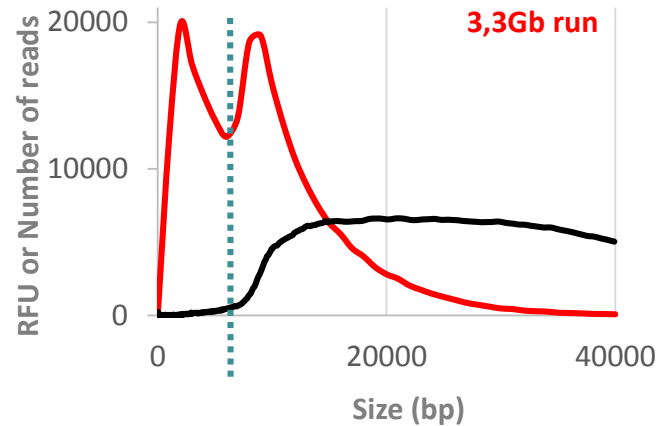
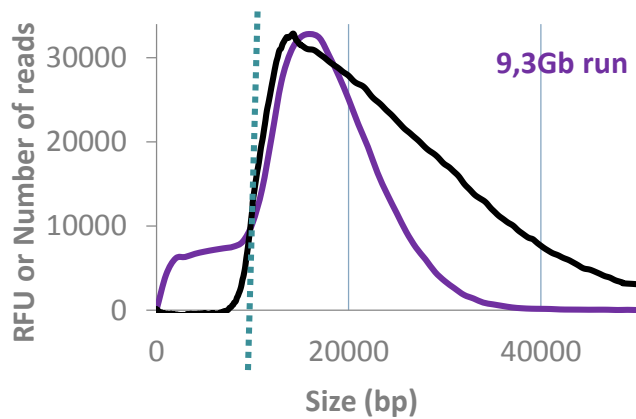
Assesment of the technology

DNA quality



Contaminant

- Less sensitive than the Pacbio
- Might have an impact (lib prep / sequencing)



Library

Reads

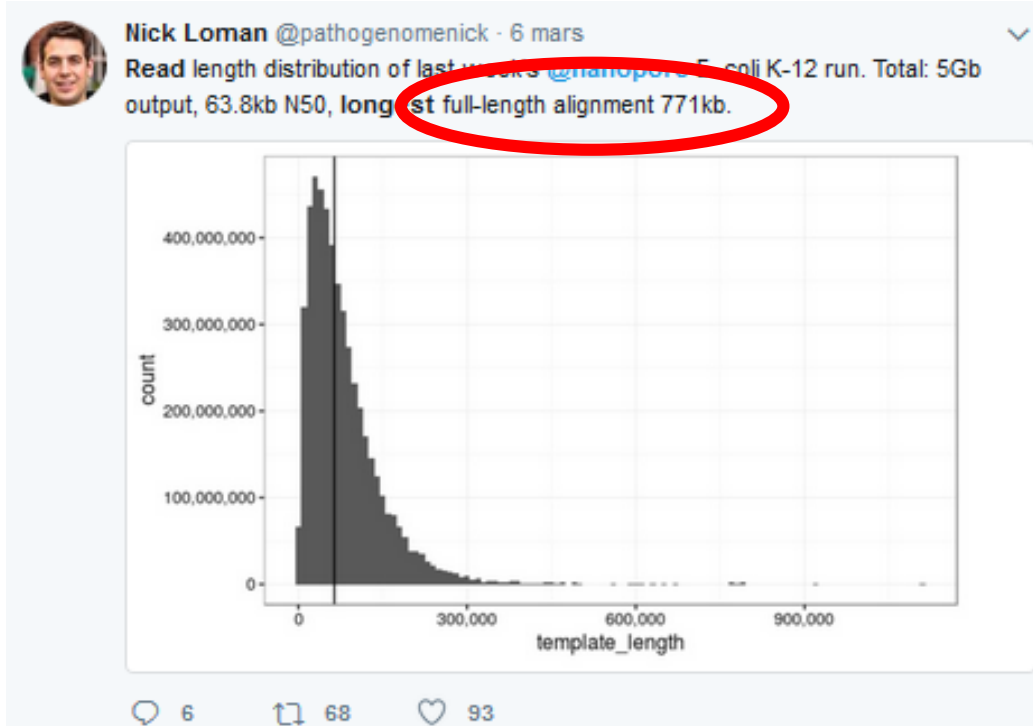
Blue Pippin

Cutoff



Assesment of the technology

DNA quality

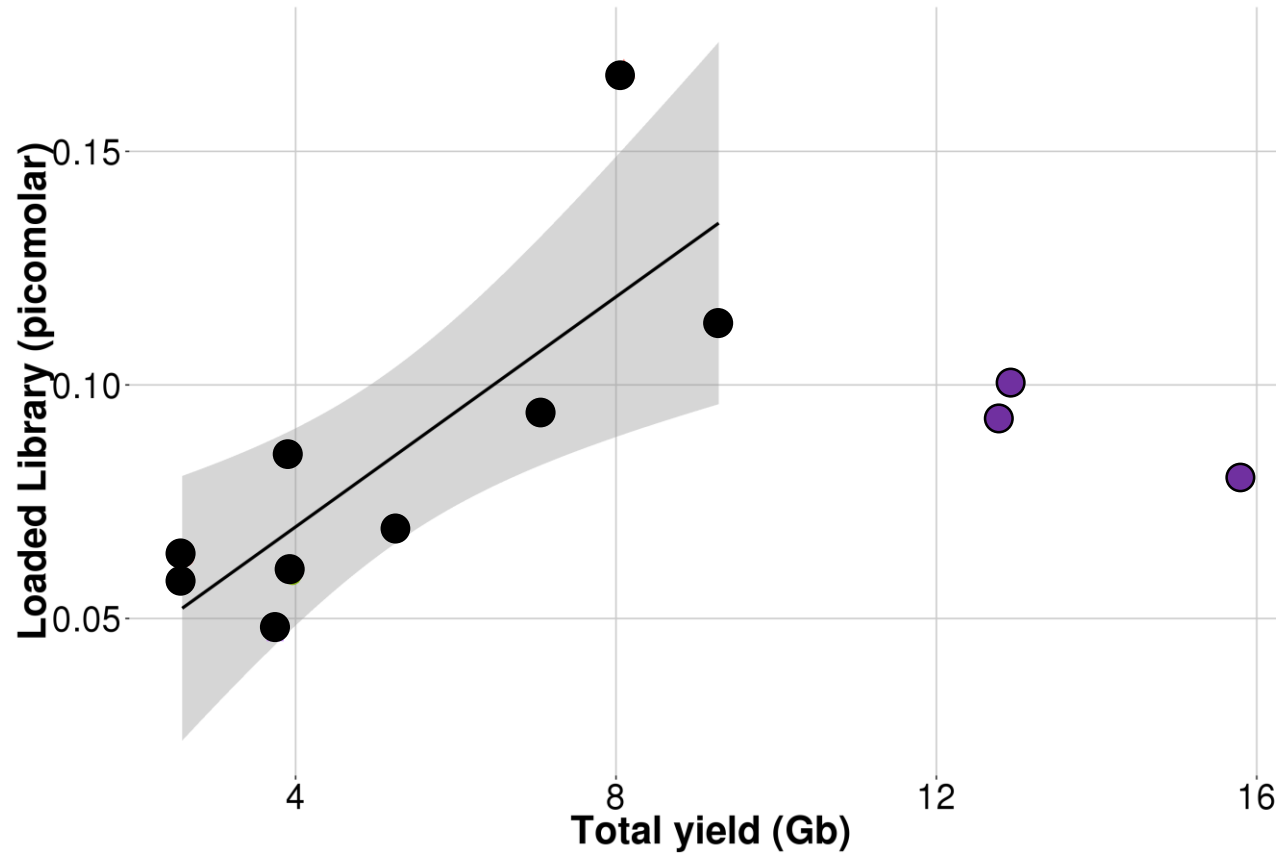


Extraction : key step

- Column
- Plugs
- Phenol/chlo

Assesment of the technology

DNA quantity



Example : 8 Gb

8 kb = 1,5 μ g

50 kb = 10 μ g

50 kb (size selection) = 20 – 30 μ g

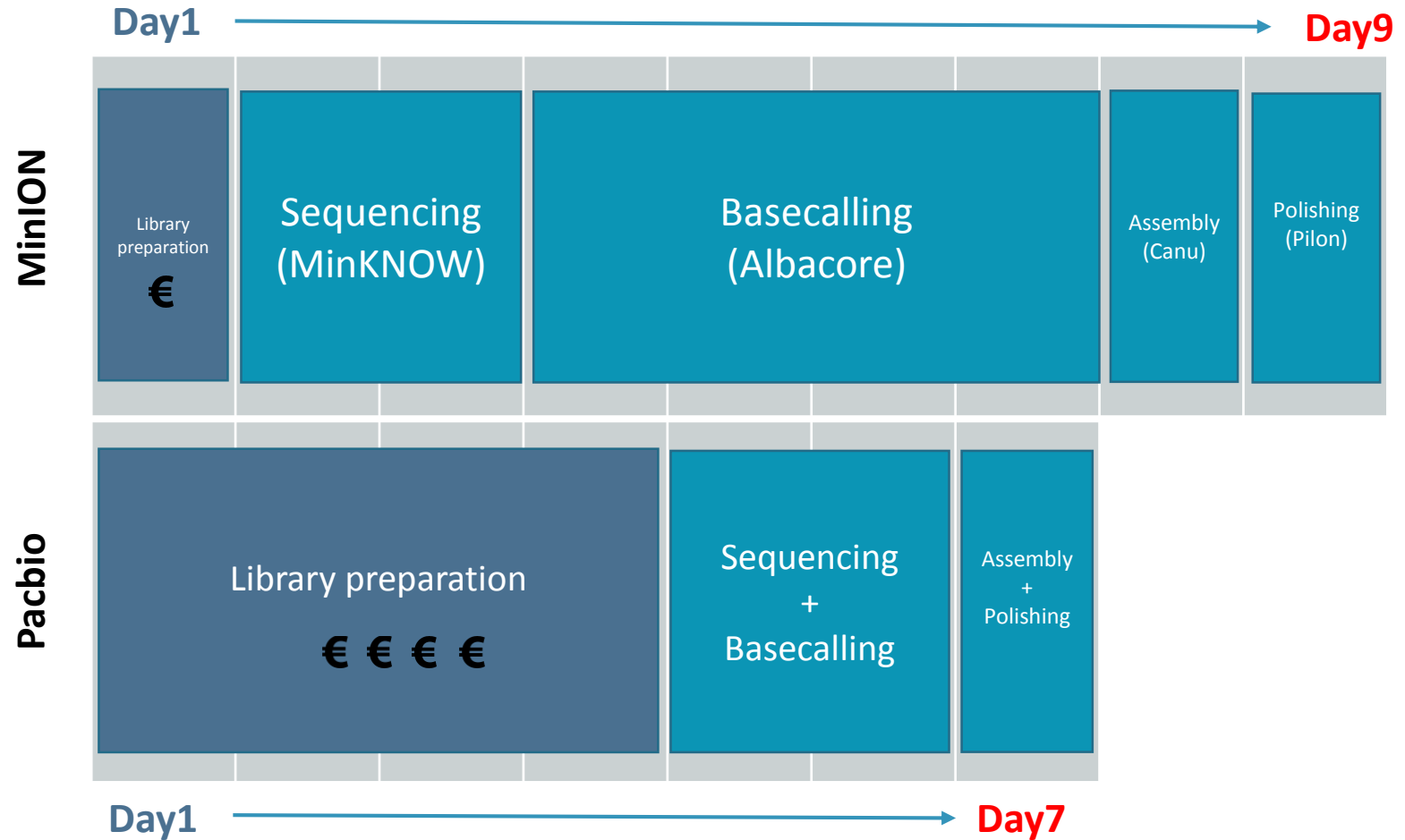
- ② **New MinKNOW version : Same amount of DNA, better yield**
- ② **Number of molecules : long reads needed \rightarrow high amount of DNA required**

Assesment of the technology

When will I get my results ?

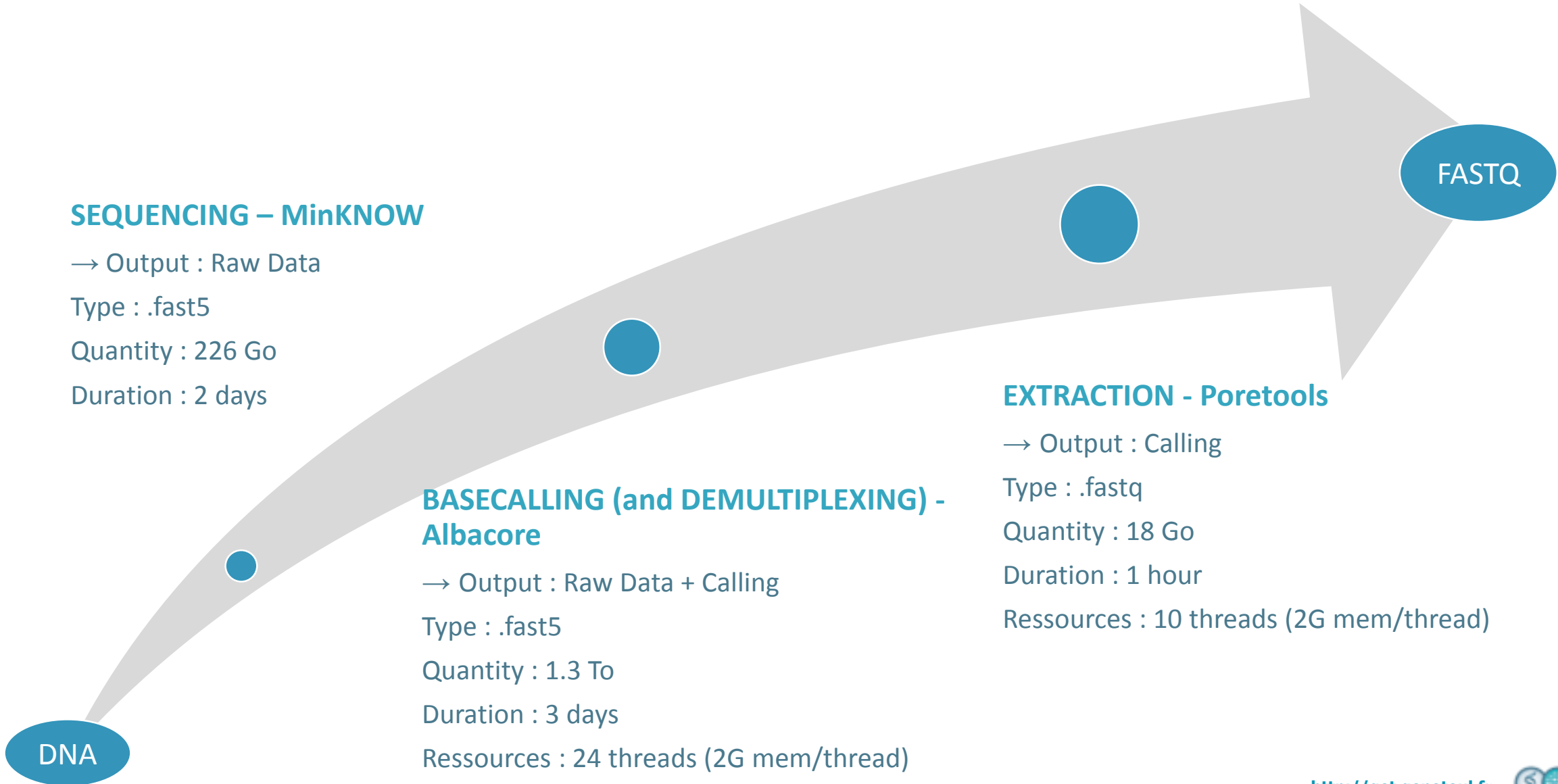
Example of a bacterial genome assembly :
Same quantity of data / Same informatics resources

- Faster & Cheaper
⚠ 1 lib = 1 FC
- More robust
- Able to run in Multiplex
- Beads (1D protocol) : limit for very high molecular weigh DNA (> 60kb)
- rapid kit / ratio transposase



Informatic pipeline 1/2

Bacterial genome : 5Mb, 1D+ protocol



SEQUENCING – MinKNOW

→ Output : Raw Data

Type : .fast5

Quantity : 226 Go

Duration : 2 days

BASECALLING (and DEMULTIPLEXING) - Albacore

→ Output : Raw Data + Calling

Type : .fast5

Quantity : 1.3 To

Duration : 3 days

Ressources : 24 threads (2G mem/thread)

EXTRACTION - Poretools

→ Output : Calling

Type : .fastq

Quantity : 18 Go

Duration : 1 hour

Ressources : 10 threads (2G mem/thread)

FASTQ



Informatic pipeline 2/2

Bacterial genome : 5Mb, 1D+ protocol

ASSEMBLING - Canu

→ Output : Assembly

Type : .fasta(q)

Quantity : 0,5 Go

Duration : 3 hours

Ressources : 8 threads (2G mem/thread)

ILLUMINA POLISHING – Pilon

→ Output : Polished.assembly

Type : .fasta(q)

Duration : 1 day

Ressources : 8 threads (2G mem/thread)

CONTIG

FASTQ

Assembly quality

Accuracy of raw data on reference genome and completeness with BUSCO

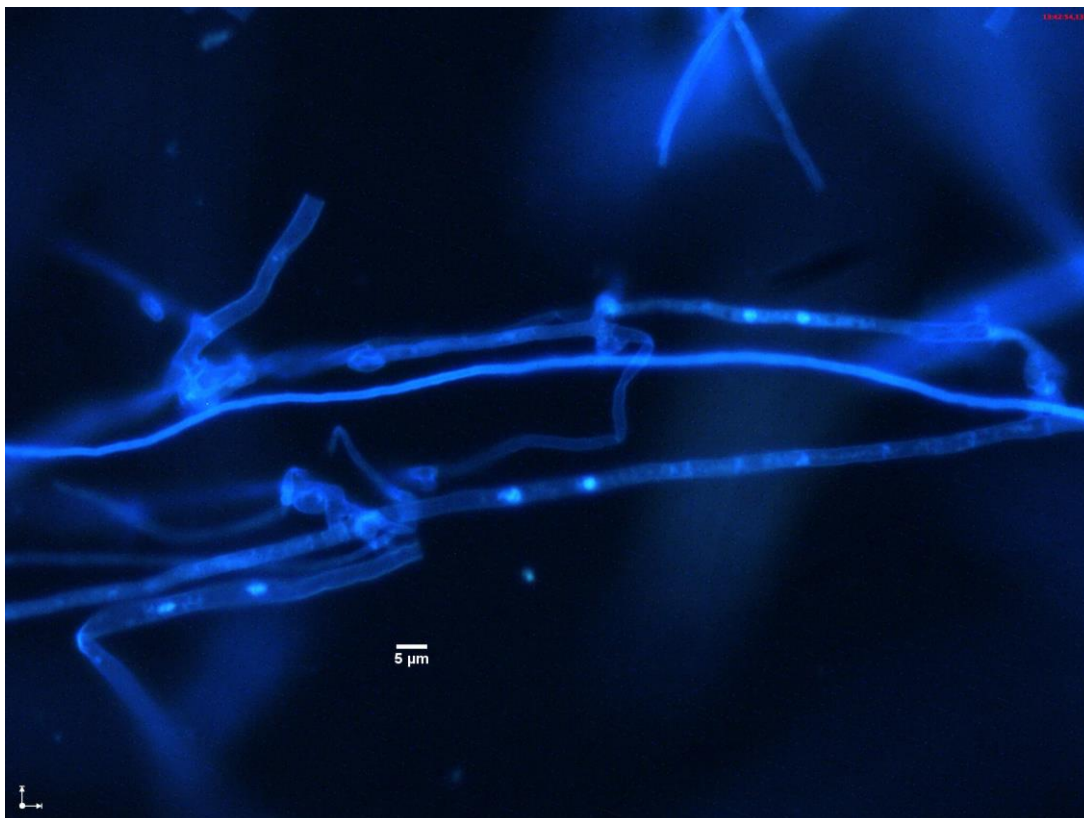
Example of a bacterial genome assembly
5Mb
Ref : Pacbio assembly (CANU)

Technology	Accuracy
Illumina	99.7 %
Pacbio	81.3 %
MinION Albacore V0	81.0 %
MinION Albacore V1	86.3 %

Assembly sets	% Complete genes	% Fragmented genes	% Missing genes
MinION-CANU	12.2	27	60.8
MinION-nanopolish	71.6	15.5	12.9
MinION-pilon	95.3	0	4.7
Pacbio-hgap3	95.3	0	4.7

Complex genome

Ganoderma boninense – Oil palm pathogen



2 nucleus
Heterozygous

Illumina results :

Number of contigs	10164
Longest contig	170 kb
N50 contig length	12 Kb
L50 contig count	1367

Repeated regions > 10 kb

50 Mb

Assembly quality

Accuracy of raw data on reference genome and completeness with BUSCO

**Exemple of a more complex
mushroom genome assembly
50Mb**

Ref : Pacbio assembly (CANU)

Assembly + Correction

Pacbio (CANU + Quiver)

Nanopore (CANU + Pilon)

~500 contigs

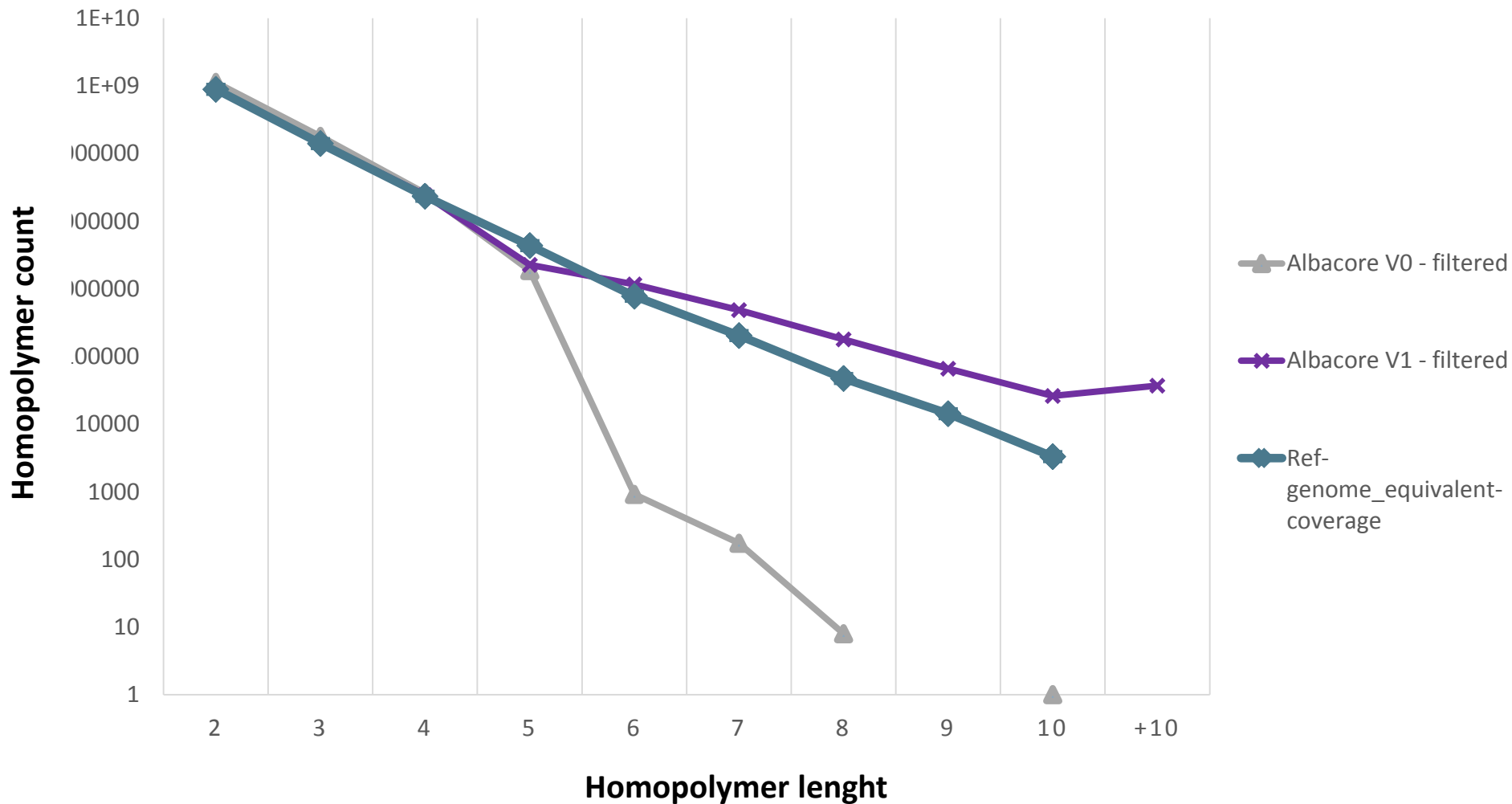
	Pacbio	Nanopore
Library prep	5 days (€€€€)	1 day (€)
Reagents	6 SMRT	1 FC
Quantity of data	4.4 Gb	4.3 Gb

WindowMaker
Haplomerger
1x Pilon

	<i>Ref + Alt</i>	<i>Ref + Alt</i>
Genes complete	92 %	89 %
Proteins found	89 %	86 %

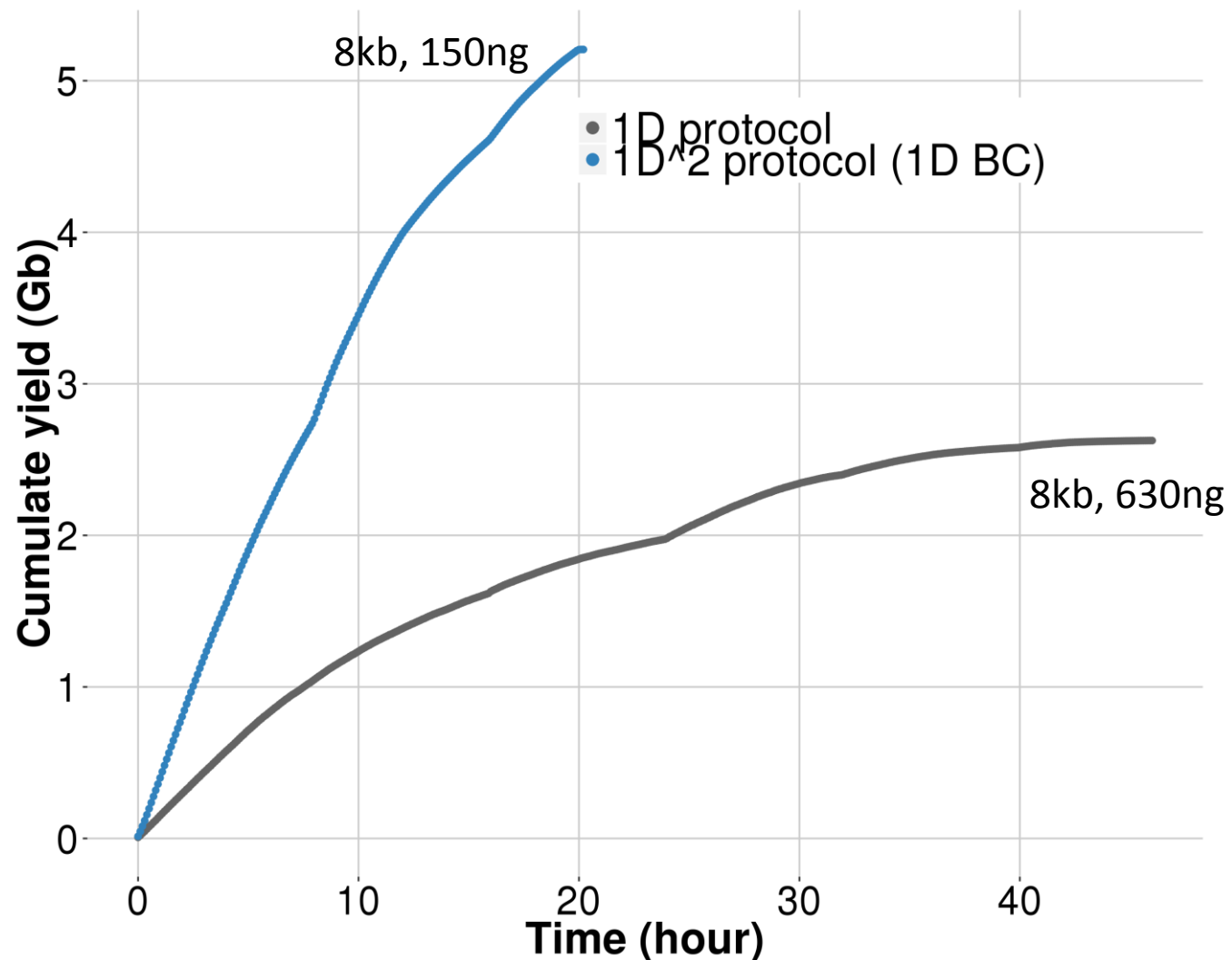
Long homopolymers issue

Is the new transducer basecaller correct this problem ?



And next ?

New chemistry : 1D² / R9.5



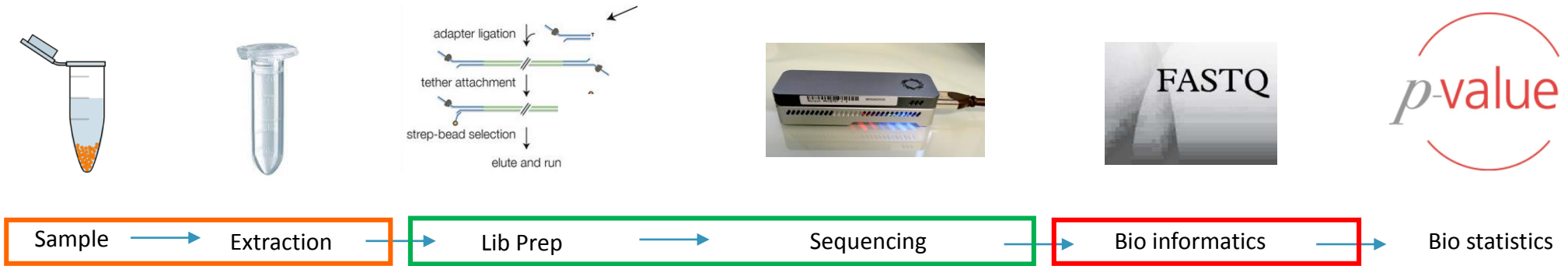
Ⓢ Basecalling 1D² :

Still running after 14 days with
Albacore v1.2.1.

ONT seems to have many
problems to correctly BC 1D²
reads

Conclusion

⑤ **MinION technology : Long DNA fragments = long reads**



⑤ **To go further and improve genomes assembly : scaffolding technologies (10X Genomics, BioNano)...**

Tomato genome



Technology	N50
PacBio (RSII 70 x)	3.2 Mb
+ Bionano (2 enzymes)	32 Mb
+ Chromium + Illumina (100x)	45 Mb

Thanks !



Céline Vandecasteele, Claire Kuchly,
Cécile Donnadiou, Olivier Bouchez, Gérald Salin,
Denis Milan, Alain Roulet, Céline Roques



Christophe Klopp, Didier Laborie, Marie-Stephane Trotard



Frédéric Breton, Alexandra Vaillant, Leatizia Camus-Kulandaivelu



Tan joon sheong, Sharifah Shahrul Rabiah Syed Alwee, Kwan Yen Yen



Baptiste Mayjonade, Jérôme Gouzy, Fabrice Roux



Guillaume Croville, Jean-luc Guerin



Caroline Callot, Stéphane Cauet, Hélène Berges



Mohammed Zouine, Pierre Frasse, Mondher Bouzayen

