2017 LONDON CALLING

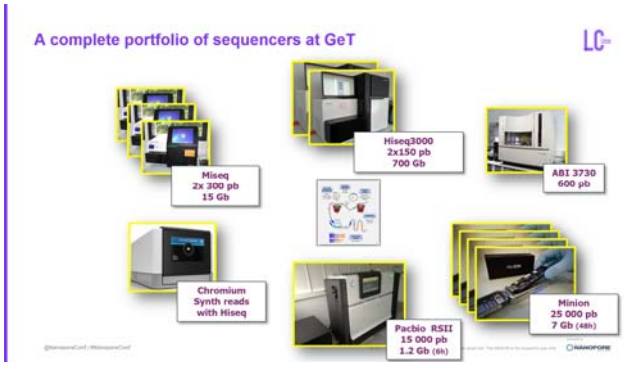
GeT experience on long fragments technologies

Denis Milan, INRA & Genotoul, Toulouse, France

http://get.genotoul.fr
@get_genotoul

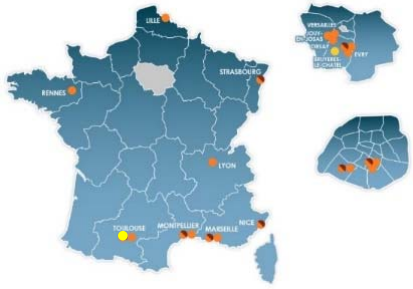
GeT Platform at Toulouse (France)

Genomics and Transcriptomics (GeT) Platform of Genotoul hosted by INRA SCIENCE & IMPACT



A strong partnership with Bioinformatics platform Genotoul Bioinfo

A node of the National Distributed Infrastructure « France Génomique » (60 M€ / 8 y) FRANCE GÉNOMIQUE



Quality certification ISO9001 & NFX 50 900, Propel Illumina PROPEL Certified 2016

ORGANISED BY Oxford NANOPORE Technologies

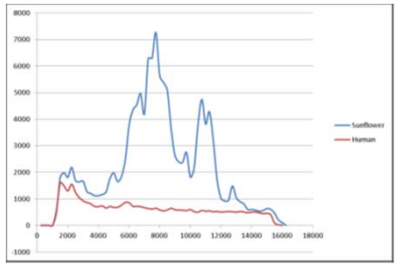
1

First experience on PacBio RSII

Evolution of Long Fragments sequencers

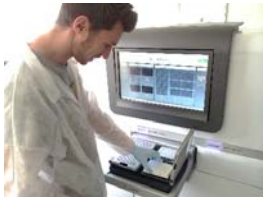
A need for long fragments sequencing to sequence plant genomes

Sunflower : 30 % of repeated sequences (LTR)
Human : 8.8 % of repeated sequences



January 2015 :

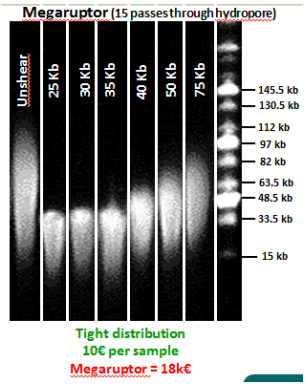
- Oxford Nanopore was ever in development
- P6/C4 chemistry of PacBio permits sequencing of longer fragments → Investment in a PacBio RSII

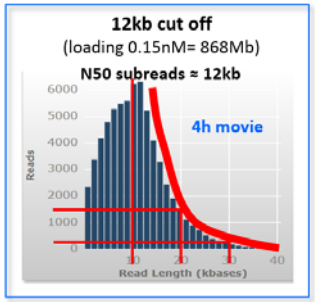


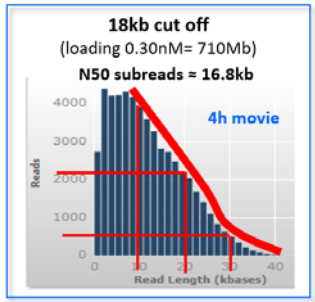
Optimization on PacBio RSII

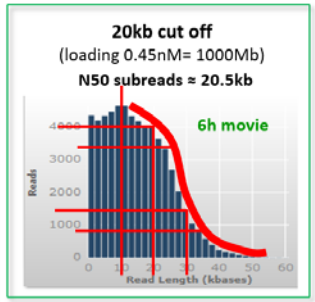
Improvement of read length :

- Shearing with Megaruptor
- Sizing with Blue Pippin
- Increase of run length





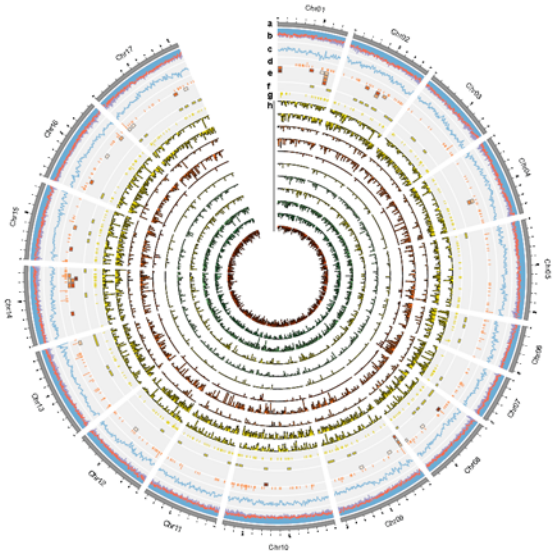




Assemblies of Sunflower genome

- INRA (Sunflower Team) :
  Hiseq 127 X
  → 43 % coverage
- International consortium :
  454, Hiseq, Genetic and physical map
  →63 % coverage
- INRA (Sunflower Team) :
  PacBio 107 X (407 SMRT)
  →84 % coverage

13124 contigs N50 = 498 kb
+90 % anchor



Longest subreads
80974 bp
79860 bp
79834 bp
78105 bp
77481 bp

Nicolas Langlade, Stephane Munoz, Jérôme Gouzy, Baptiste Mayjonade,

2

Test of Minion potential

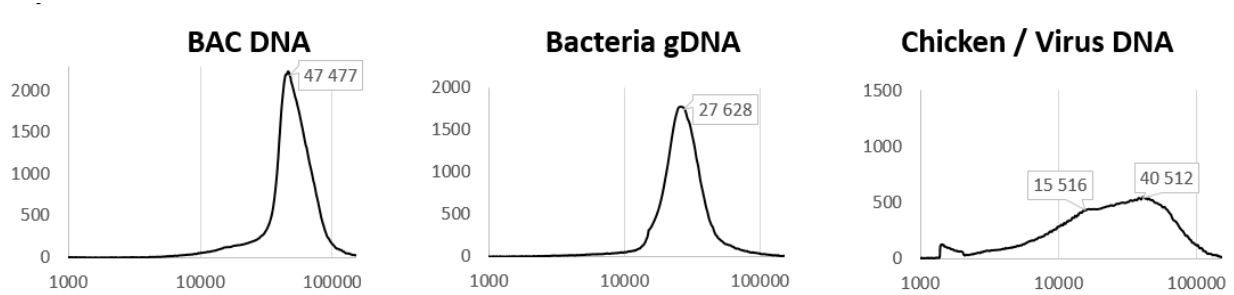
Back to the future …

Minion pilot projects

Setting up pilot projects :

- 3 kind of materials : BAC Clones, bacterial DNA, Virus infecting animals
- 3 kind of kits : Rapid, 1D, 2D
- 3 kind of basecallers : Minknow (local), Metrichor (cloud), Albacore (local)

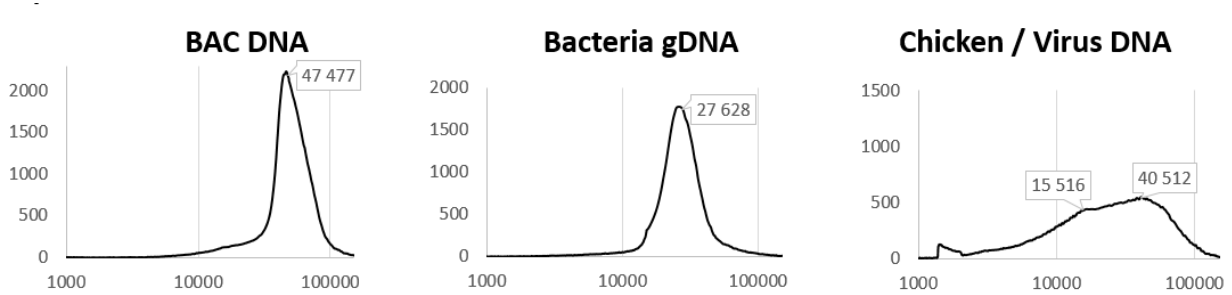
First results on Rapid Run



Preparation kit used and Amount of DNA required	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)
Rapid 200 ng	0,1	37,9	0,5	10,1	0,3	33,5

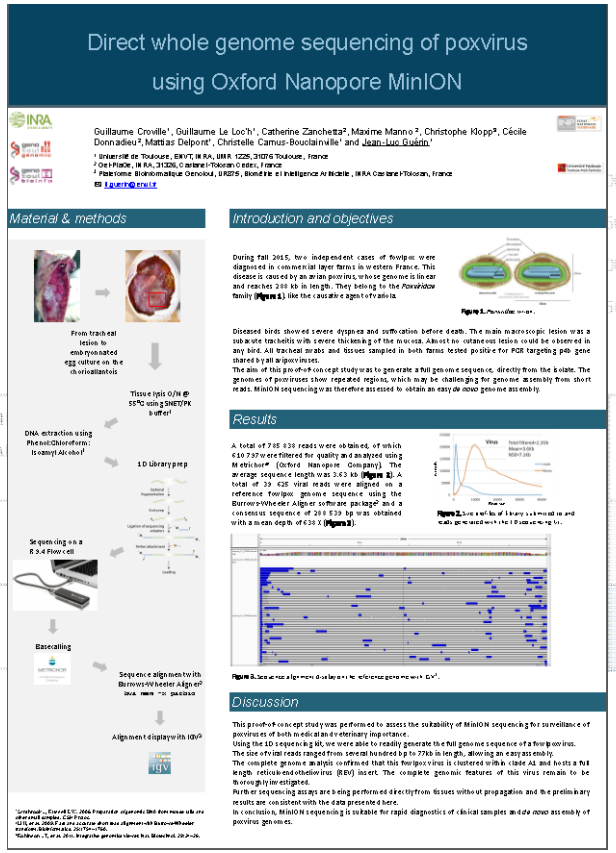
First results on 1D run

Poster 22



Preparation kit used and Amount of DNA required	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)
Rapid 200 ng	0,1	37,9	0,5	10,1	0,3	33,5
1D 1,5µg	2,6	9,6	4,0	8,7	3,9	9,4

First results on 1D run

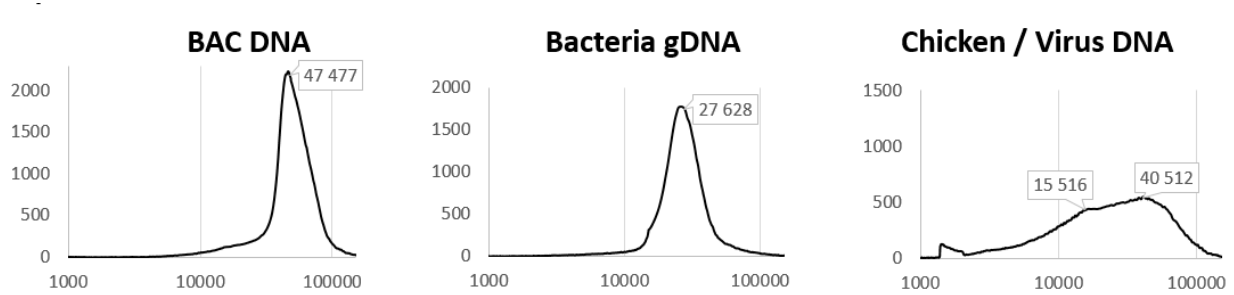


Poster 22

Guillaume Croville, Guillaume Le Loc'h, **Catherine Zanchetta**, Maxime Manno, Christophe Klopp, **Cécile Donnadieu**, Mattias Delpont, Christelle Camus-Bouclainville and **Jean-Luc Guérin**

A total of 785 838 reads were obtained, of which **610 797** were filtered for quality and analyzed using Metrichor® (Oxford Nanopore Company). The size of viral reads ranged from several hundred bp **to 77kb** in length. The average sequence length was **3.63kb** (Figure2). A total of 39 625 viral reads were aligned on a reference fowlpox genome sequence using the Burrows-Wheeler Aligner software package2 and a **consensus sequence** of **288 539bp** was obtained with a mean depth of **638X** (Figure3).

First results on 1D run

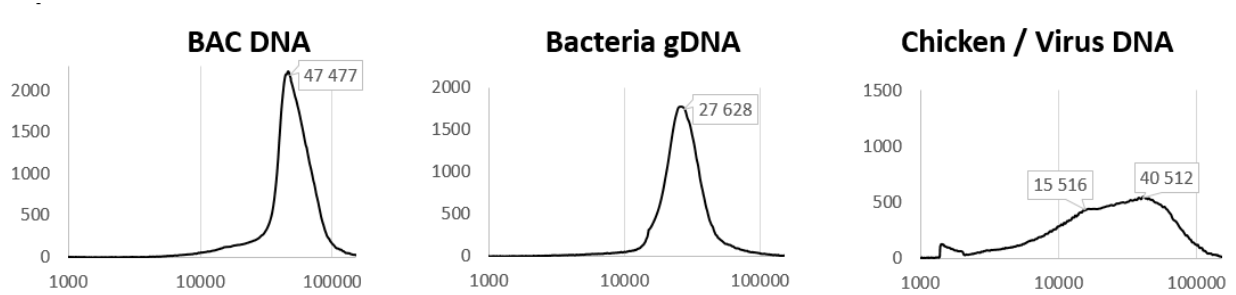


Preparation kit used and Amount of DNA required	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)
Rapid 200 ng	0,1	37,9	0,5	10,1	0,3	33,5
1D 1,5µg	2,6	9,6	4,0	8,7	3,9	9,4

Shearing with Megaruptor (20 kb)

Sizing with Blue Pippin (11-50 kb)

Improvement of 1D to “1D+”

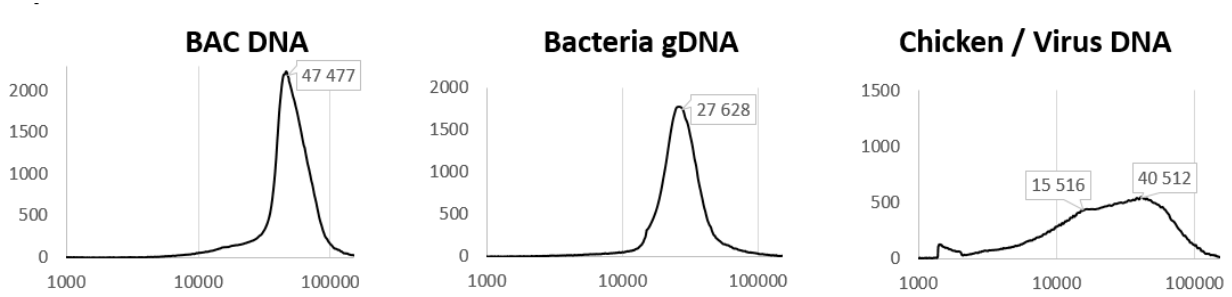


Preparation kit used and Amount of DNA required	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)
Rapid 200 ng	0,1	37,9	0,5	10,1	0,3	33,5
1D 1,5µg	2,6	9,6	4,0	8,7	3,9	9,4
1D+ $\frac{1{,}5 * mean\ size}{8}$			9,3	19,0	3,3	15,1

Shearing with Megaruptor (20 kb)

Sizing with Blue Pippin (11-50 kb)

Improvement of 1D to “1D+” to “1D++”

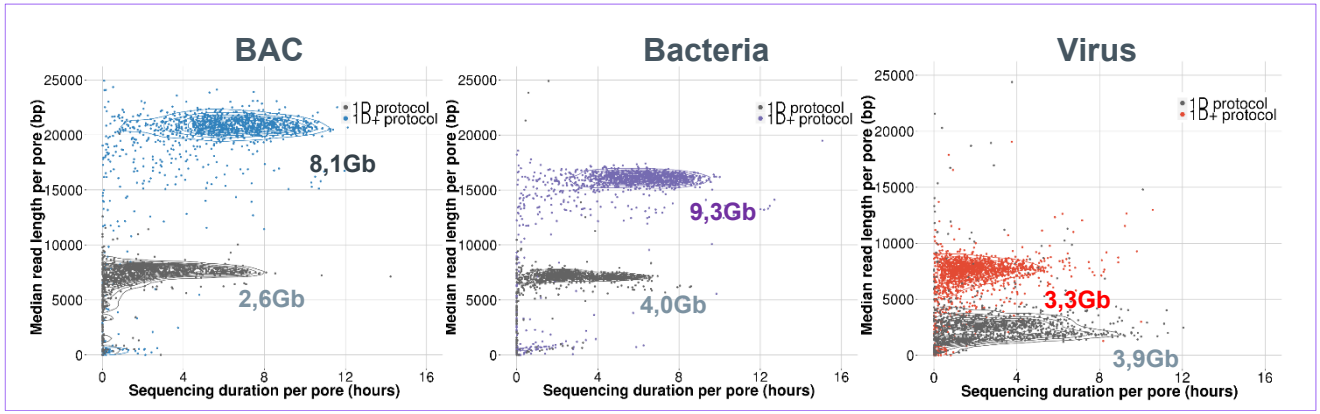


Preparation kit used and Amount of DNA required	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)	Total Filtered (Gb)	N50 (Kb)
Rapid 200 ng	0,1	37,9	0,5	10,1	0,3	33,5
1D 1,5µg	2,6	9,6	4,0	8,7	3,9	9,4
1D+/ ++ $\frac{1{,}5 * mean\ size}{8}$			9,3	19,0	3,3	15,1

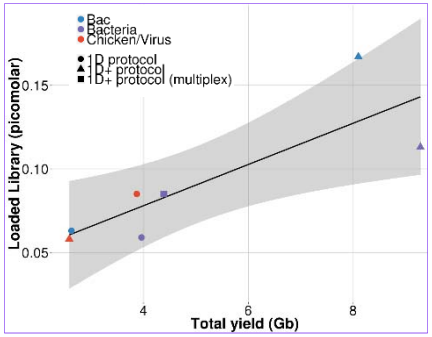
Shearing with Megaruptor (40 kb)

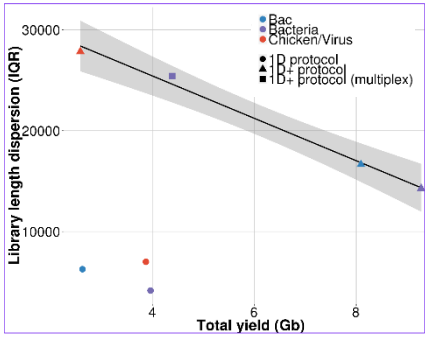
Sizing with Blue Pippin (16-50 kb)

Some preliminary feelings to discuss



No overloading

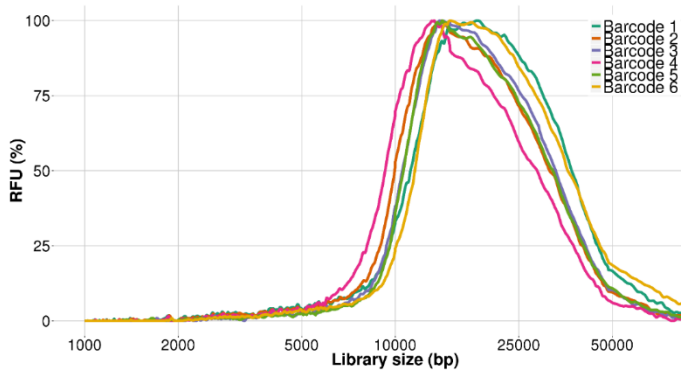




Yield might be correlated to the dispersion of Fragment size

Efficiency of Barcoding (January version of the kit)

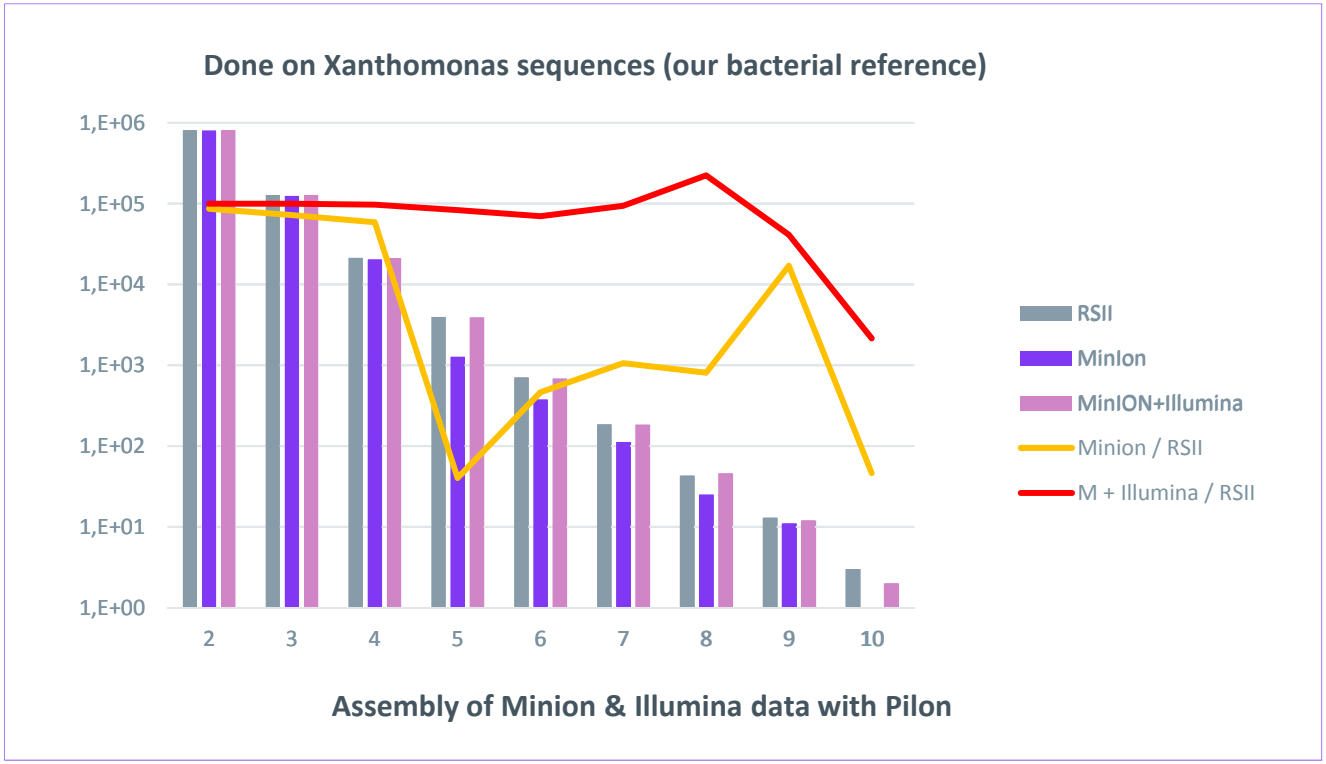
Library profile for each Barcode



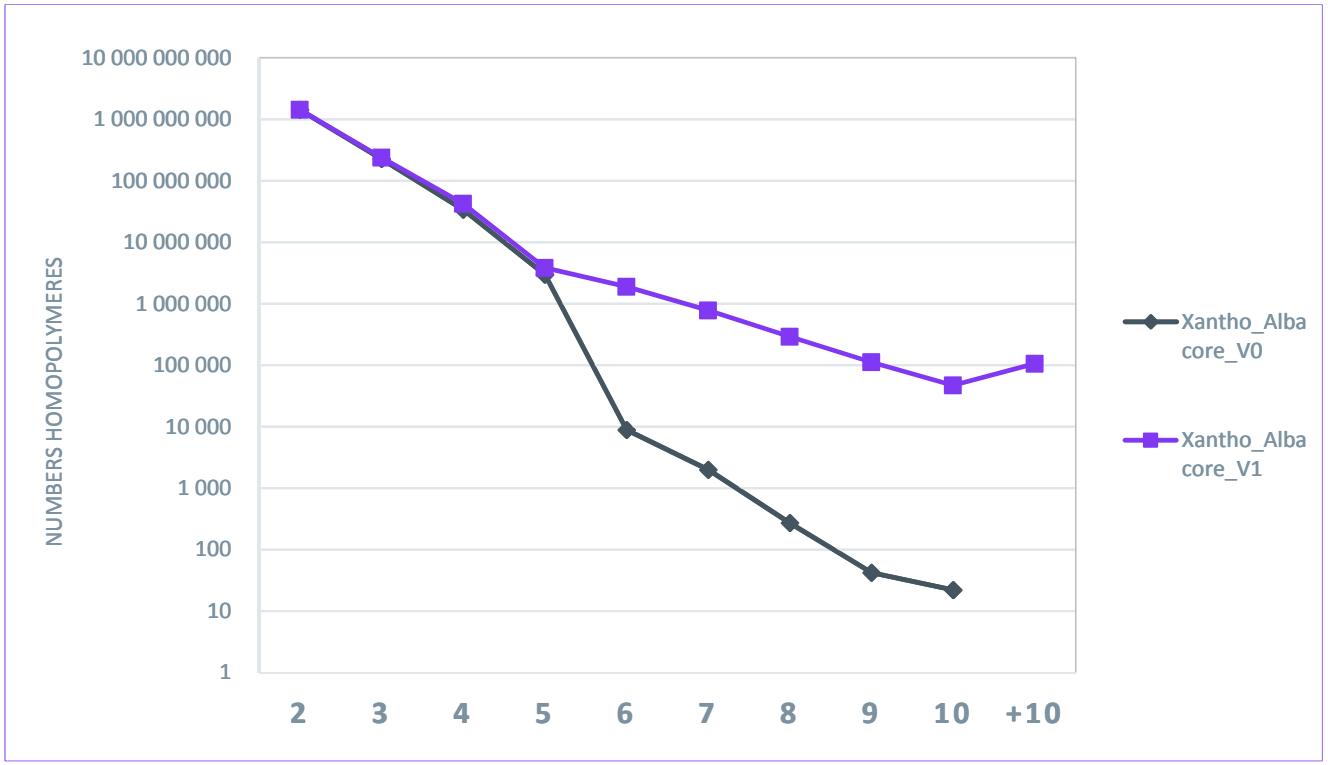
	% of reads per barcode	N50 (Kb)
1	17	10
2	17	13
3	13	10
4	21	16
5	15	11
6	17	13

Barcoding is efficient (at least at that level) with **homogenous** results

Detection of homopolymeres



Improvement of Homopolymere sequencing with Albacore v1



Analysis of sequence accuracy on Xanthomonas data set

BUSCO2 : Identification of 148 conserved genes in assembly genome of E Coli

	% Complete genes	% Fragmented genes	% Missing genes
X_pacbio_hgap3	95 %	0 %	5 %
X-Av1-F10_minion_canu	17 %	36 %	47 %
X-Av1-F10_minon_canu_Illumina_pilon	95 %	0 %	5 %
X-1D_Metrichor_minion_canu_nanopolish	72 %	15 %	13 %

Poster 6

Sequencing of a local strain of Arabidopsis Thaliana

Sequencing on MinION at 80 x
90 % of the sequence in 20 contigs

NUM	187
MIN	5257
MAX	15 602 179
N50 BP	7 697 404
N50 NUM	6
N90 BP	1 036 635
N90 NUM	20
MEAN	644 576
MEDIAN	35 272
BP	120 535 815

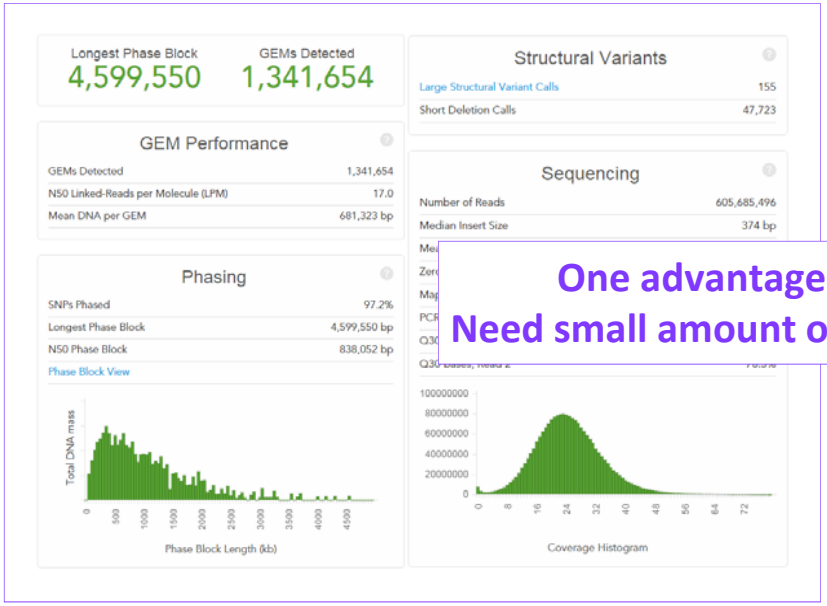
Baptiste Mayjonade, Fabrice Roux, Jérome Gouzy

3

Other technologies

Contribution of Chromium from 10Xgenomics to genome sequencing

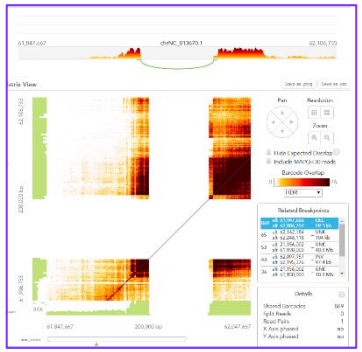
Pilot projects on Rabbit for phasing



One advantage
Need small amount of DNA

Longest phase block 4.6 Mb

20 % of representation



Julie Demars
Christophe Klopp

Structural Variants detection

Assembly of Tomato genome using combination of technologies

Chromium (10 X Genomics)

Irys (Bionano)

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

Pilot projects on Tomato for de novo assembly

Technology	N50
Chromium + Illumina (100x)	1.8 Mb
+ Bionano (2 enzymes)	17 Mb

GBF Lab : Mohammed Zouine, Pierre Frasse, Mondher Bouzayen
CNRGV lab (Bionano) : Sandrine Arribat, William Marrande, Hélène Bergès

4

Conclusions

Conclusions

Oxford Nanopore has a great potential

- **Very long** reads
- A simple possibility to provide the technology at **various scales** (from SmigdION to PromethION)
- A potential to distinguish **modified bases**
- A potential to challenge Illumina for the **Cost / Gb**

Different technologies available for long reads

- **PacBio** : A golden standard up to now for long reads (no bias in sequence errors)
- **10x** : Low amount of DNA, Illumina accuracy, low cost in complement of Illumina
- **ONT** : A greater potential of evolution, but up to now a less mature technology

Key points

- **Ultra long DNA** is required, when possible in large amount
- The infrastructure for the **analysis of data** should not be underestimated
- An interest to **combine different technologies** depending on applications, requested specifications & money available

Thanks to …

INRA GeT Platform : Cécile Donnadieu, Catherine Zanchetta, Pauline Heuillard, Olivier Bouchez *et al*
Maxime Manno, Claire Kuchly, Céline Vandecasteele

INRA Bioinfo platform : Christophe Klopp, Christine Gaspin

INRA CNRGV : Sandrine Arribat, William Marrande, Caroline Callot,
Stéphane Cauet, Hélène Bergès

INRA/CNRS LIPM : Baptiste Mayjonade, Stephane Munoz, Nicolas Anglade,
Fabrice Roux, Jérome Gouzy

INRA/ENSAT : GBF : Mohammed Zouine, Pierre Frasse, Mondher Bouzayen

INRA GenPhySE : Julie Demars, Isabelle Hochu

ENV/INRA : Guillaume Croville, Guillaume Le Loc'h, Mattias Delpont, and Jean-Luc Guérin

Present at London Calling

- Cécile Donnadieu (Head of GeT) **Poster 6**

- Catherine Zanchetta (Minion Developments) **Poster 6**

- Jean-Luc Guérin (PI Virology project) **Poster 22**

2017
LONDON
CALLING

ORGANISED BY