



cgfb

GÉNOME TRANSCRIPTOME

03 octobre 2018

# minION- Application

*Franck Salin, Christophe Boury, Charlotte Mouden*

Colloque EPGV  
3-5  
Octobre 2018

Which nanopore device is best for you?



**Flongle**                      **MinION**                      **GridION**                      **PromethION**

Long read, direct DNA/RNA/epigenetic sequencing, scalable, real time/rapid, on-demand sequencing that is easy to use and install.

## Full comparison table

	Flongle	MinION Mk 1B	GridION X5	PromethION (1 flow cell)	PromethION (48 flow cells)
Read length		fragment length = read length. Longest read now >2 Mb			
Run time	1 min - 16 hrs	1 min - 48 hrs	1 min - 48 hrs	1 min - 64 hrs	1 min - 64 hrs
<b>Yield DNA sequencing</b>					
Theoretical maximum 1D Yield	Up to 3.3 Gb	Up to 40 Gb	Up to 200 Gb	Up to 315 Gb	Up to 15 Tb
Current yield range 1D	Early access to start ASAP - commercial target 1 Gb	Up to 30Gb (Rev D Chip)	Up to 150Gb (RevD Chip)	Up to 150 Gb	

# Illumina vs. Nanopore

## Illumina

### Reads

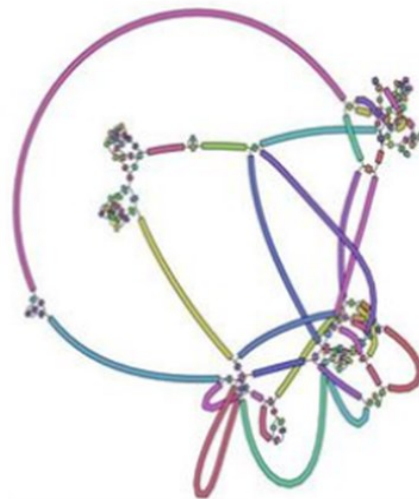
- 100–250 bp reads, 300–1000 bp fragments (shorter than repeats)
- Very accurate

### Assemblies

- Fragmented
- Small N50: 10s–100s of kbp
- Very accurate

### Uses

- SNPs
- Phylogenetics
- Specific alleles



## MinION

### Reads

- Wide length distribution, 20+ kbp common (longer than repeats)
- 90–95% accuracy

### Assemblies

- Complete
- 98+% accuracy

### Uses

- Large-scale structure
- Horizontal gene transfer



# Hybrid read sets

## ILLUMINA

**AND**

## NANOPORE

### Reads

- 100–250 bp reads, 300–1000 bp fragments (shorter than repeats)
- Very accurate

### Assemblies

- Fragmented
- Small N50: 10s–100s of kbp
- Very accurate

### Uses

- SNPs
- Phylogenetics
- Specific alleles

### Reads

- Wide length distribution, 20+ kbp common (longer than repeats)
- 80–95% accuracy

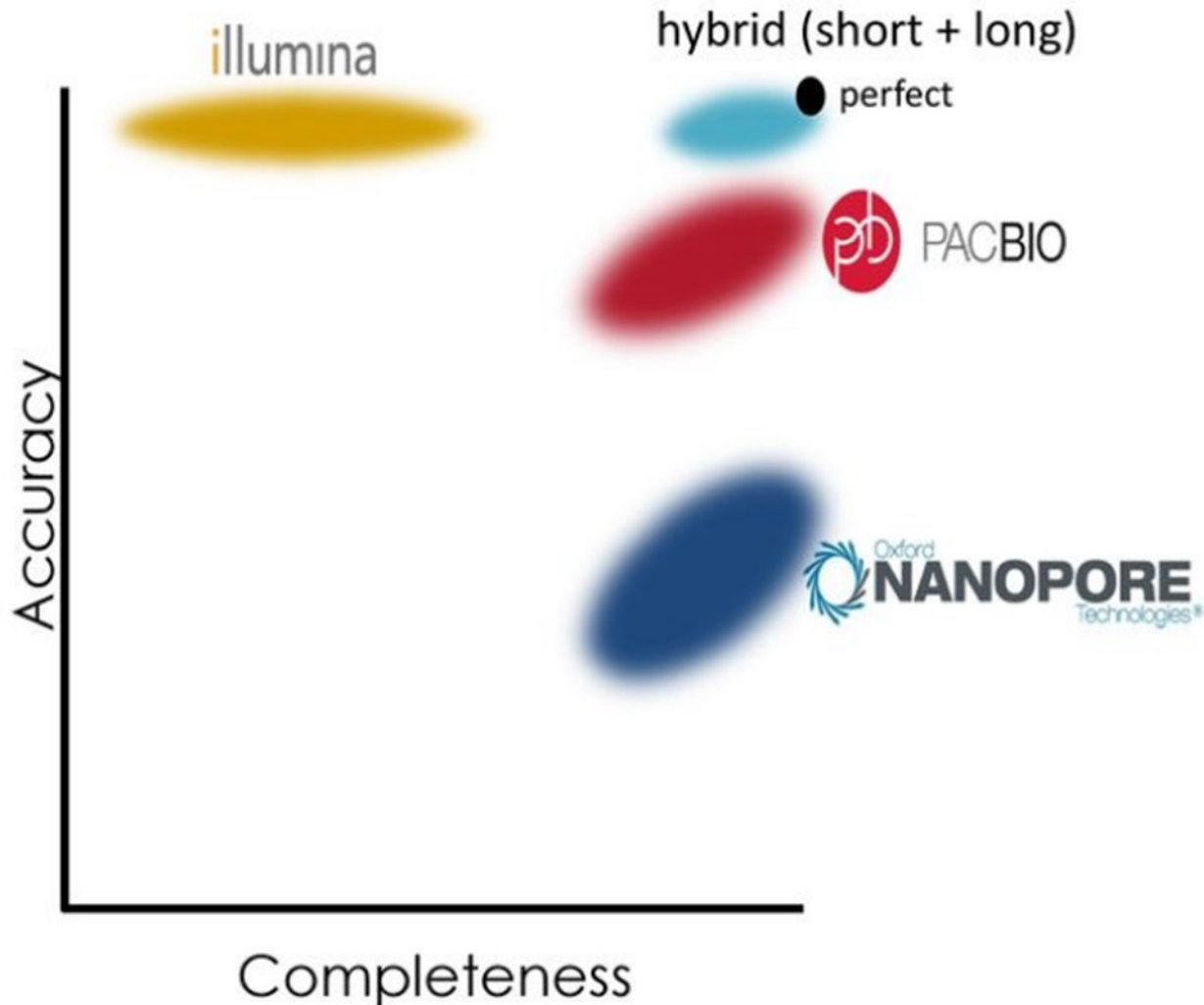
### Assemblies

- Complete
- 98+% accuracy

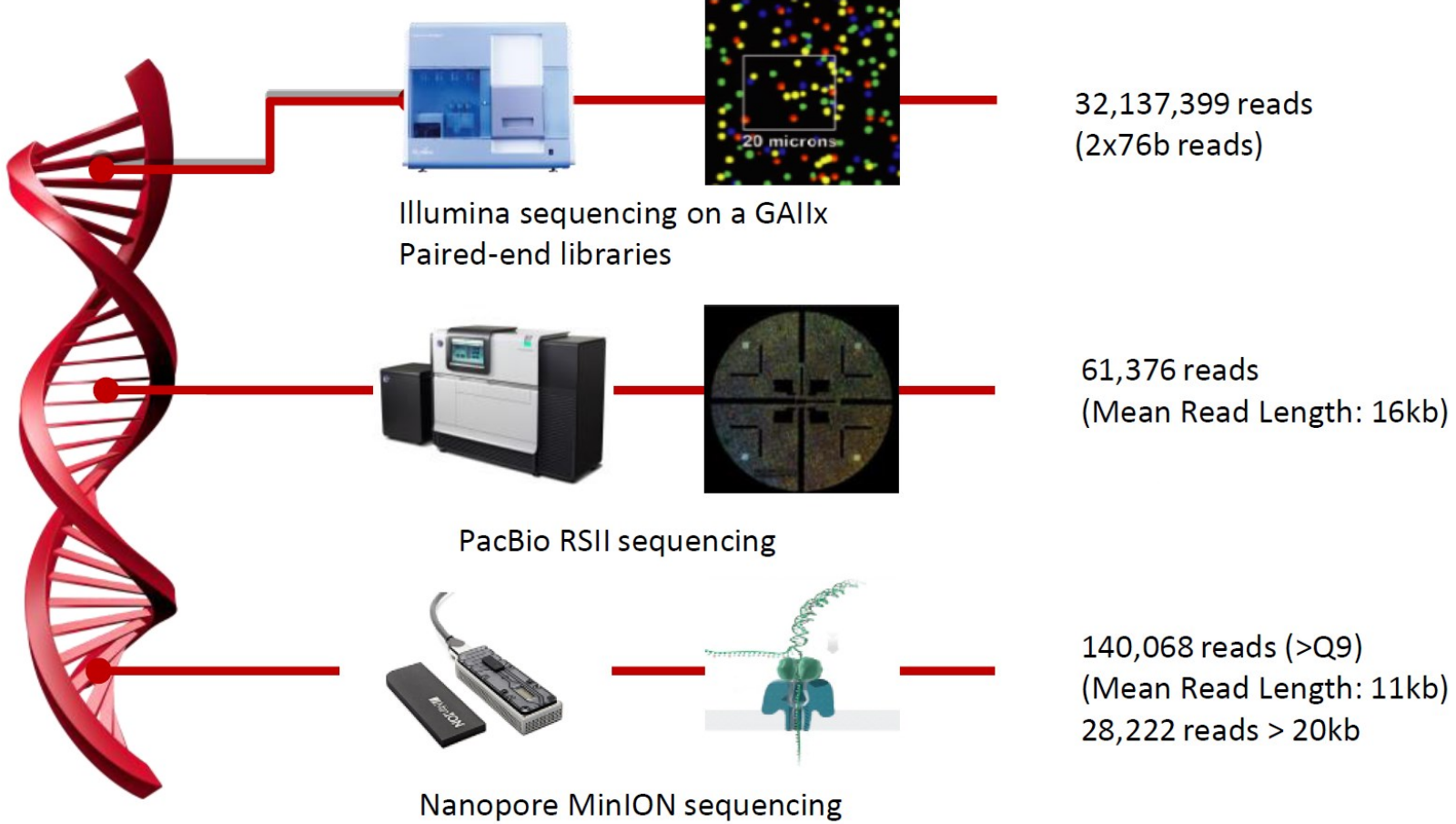
### Uses

- Large-scale structure
- Horizontal gene transfer

# Summary



### Update du génome de *Mycoplasma feriruminatoris*

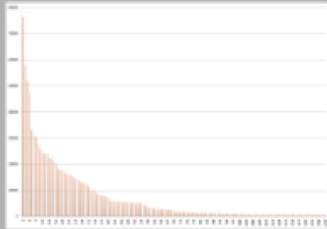


The genome of *M. feriruminatoris* sp. nov. G5847 was sequenced using Illumina and PacBio sequencing technologies at Lausanne Genomic Technologies Facility, CIG, Univ. Lausanne, Lausanne, Switzerland. ONT sequencing on MinION was achieved at the Bordeaux Sequencing facility (PGTB).

## Update du génome de *Mycoplasma feriruminatoris*

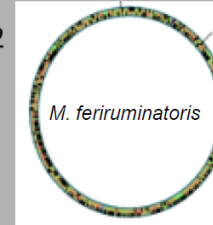
Assemblies

### Illumina



- Pipeline : Velvet 1.1
- 88 contigs
- sum : 1,017,042 bp
- average : 11,557 bp
- max : 86,939 bp
- N50 : 32,956 bp

### PacBio



- Pipeline : RS\_HGAP\_Assembly.2
- 1 circular contig
- size : 1,075,080 bp
- average coverage: 890X

### MinION + Illumina

- Pipeline : Canu1.6 + Pilon (polishing)
- 1 circular contig
- size : 1,132,484 bp
- average coverage: 1400X



### Long reads >8000kb et >Q9 ( 9.4)

- 69,808 reads
- Longueur moyenne 19,932bp
- Couverture théorique >1000X

### Short reads

- 61,376 reads
- 2x76bp reads

**Identité  
(vs short-reads assembly)**

Canu1.7

99,27 %

Nanopolish

99,54 %

Pilon1.22 x3

~100 %

Manual  
corrections

100 %

Prokka1.12

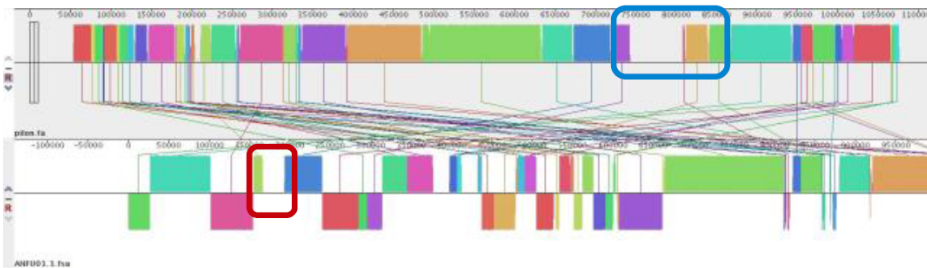
Précédente version (reads Illumina, assemblage Velvet)	Nouvelle version
88 contigs	1 contig circulaire
1,010,677 bp N50 32,956 bp	1,075,529 bp
907 gènes	948 gènes

## Update du génome de *Mycoplasma feriruminatoris*

### MinION-based assembly vs PacBio-based assembly :

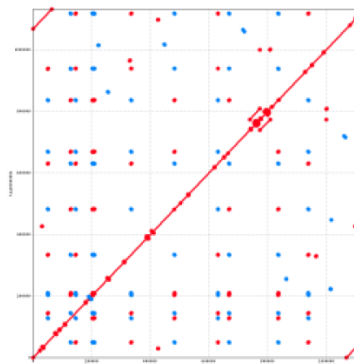
- > 100 % of the PacBio assembly matches the MinION assembly
- > MinION assembly is a bit longer

### MinION/Illumina-based hybrid assembly vs Illumina-based assembly



All Illumina contigs map with the MinION contig, except 1 (red circled), corresponding to a duplicated region in the long read assembly (blue circled).

## Results



Self-alignment of the MinION/Illumina-based hybrid assembly using showing a 35kb tandem duplication

### Gene content of the 35kb duplicated region



- > Region includes the recently described MIB-MIP genes
- > Shown to encode an Ig-binding and cleavage system (Arfi et al, 2016) and a mycoplasma specific ATPase derived from the F1FO ATP synthase (Beven et al, 2012).
- > Both elements might be involved in the capacity of the mycoplasma to escape the host immune system.

#### References

- Arfi et al. MIB-MIP is a mycoplasma system that captures and cleaves immunoglobulin G. *Proc Natl Acad Sci USA*. 2016.
- Beven et al. Specific evolution of F1-like ATPases in mycoplasmas. *PLoS One*. 2012.
- Kurtz et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004.
- Darling et al. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One*. 2010.
- Uchiyama et al. MBGD. Microbial Genome Database for Comparative Analysis. *Nucleic Acids Res*. 2003.

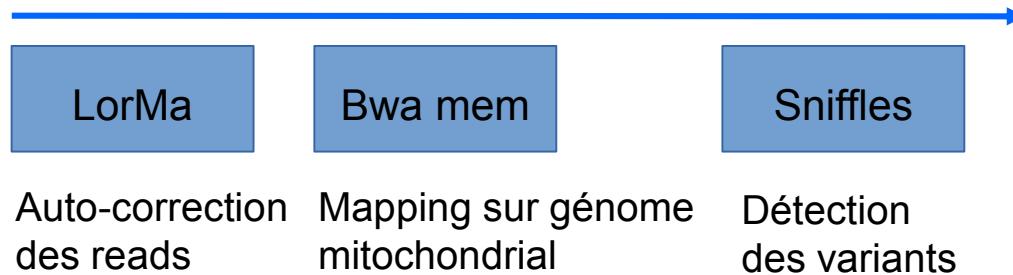
- ✓ **Ingénierie génétique en cours basée sur cette nouvelle version de génome, en vue de la production d'un vaccin**
- ✓ **Les mutants issus de l'ingénierie seront vérifiés par assemblage *de novo***



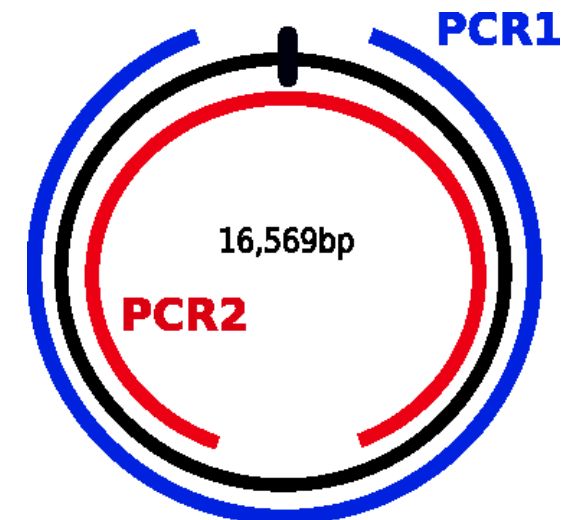
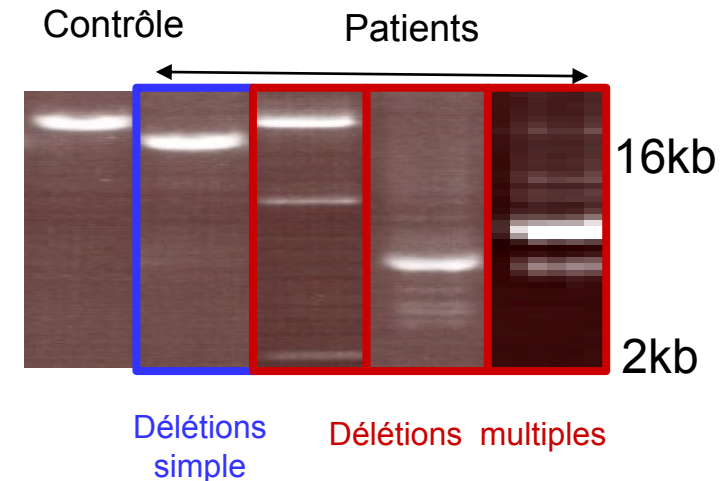
## Caractérisation de grandes délétions du génome mitochondrial humain

- Délétions complexes, pouvant être multiples au sein d'un même individu
- Responsables de pathologies mitochondriales
- Diagnostic basé sur électrophorèse de PCR long-range et NGS (PGM), mais points de cassures difficiles à identifier précisément

→ Séquençage des long range PCR en long read, grande profondeur (>4000X)



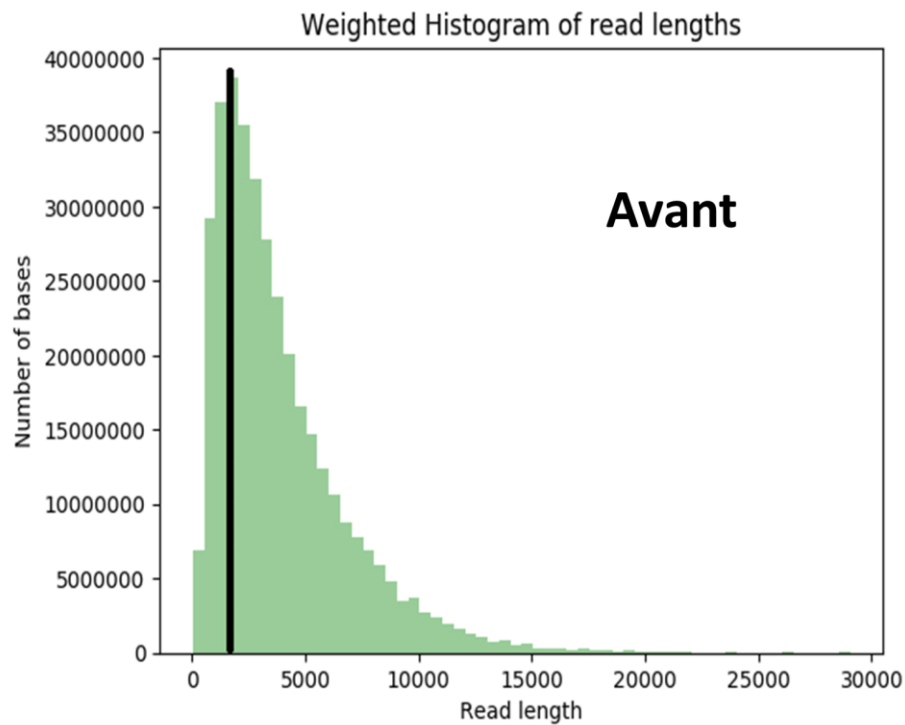
- ✓ **7 points de cassures identifiés, 4 vérifiés par Sanger, les autres en cours**
- ✓ **Malgré le taux d'erreurs élevés, des haplogroupes ont peut être correctement identifiés à partir de SNPs, chez tous les individus passés (HaploGrep)**



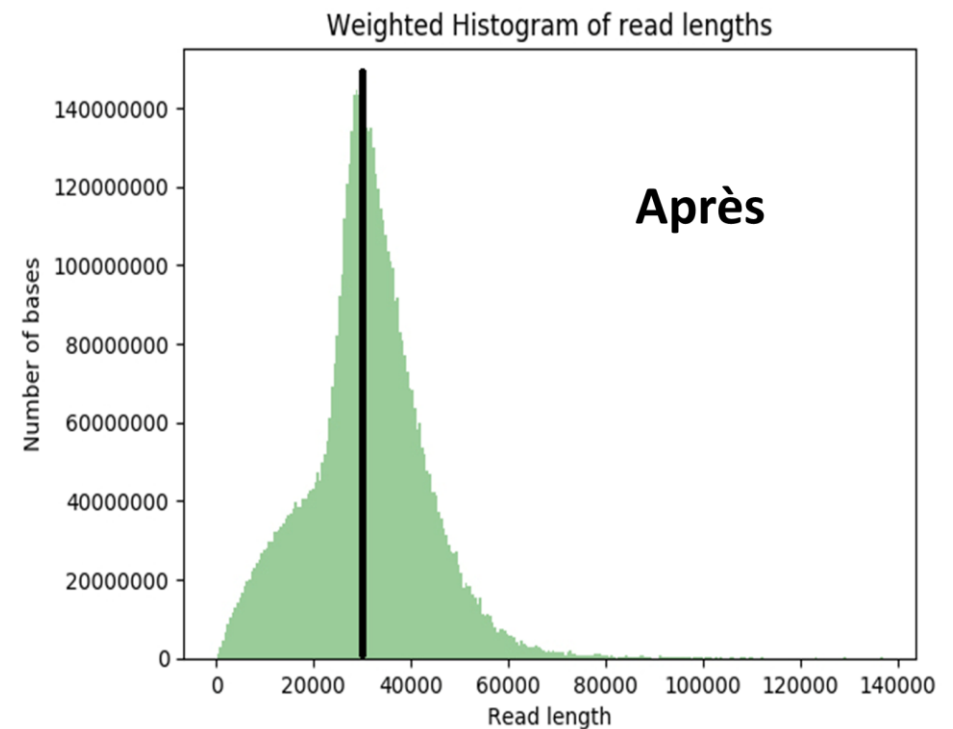
# Plateforme Génome-Transcriptome de Bordeaux

Ci-dessous un exemple illustratif de la différence de données de séquençage obtenues sur la PGTB avant et après l'optimisation d'un protocole d'extraction d'ADN (élimination des polysaccharides). Ce travail d'optimisation d'extraction de l'ADN a été effectué par Arthur Demené, doctorant de l'UMR Biogeco, sur le champignon pathogène du genre *Castanea* (*Cryphonectria parasitica*). Entre 15 et 20 fois plus de données ont été générées à partir de reads 10 à 15 fois plus longs.

**Weighted Histogram of read lengths**



**Weighted Histogram of read lengths**



	Avant optimisation d'extraction d'ADN	Après optimisation d'extraction d'ADN
Gb générées	0,36	6,2
N50 (Kb)	3,1	31 (270 000 reads)
N90 (Kb)	1,04	14,9
Médiane de longueur de reads (Kb)	1,5	25
Quantité de reads au-delà de 50 Kb	0	6169

## NEW ASIC REVISION

MinION and GridION “Rev D”

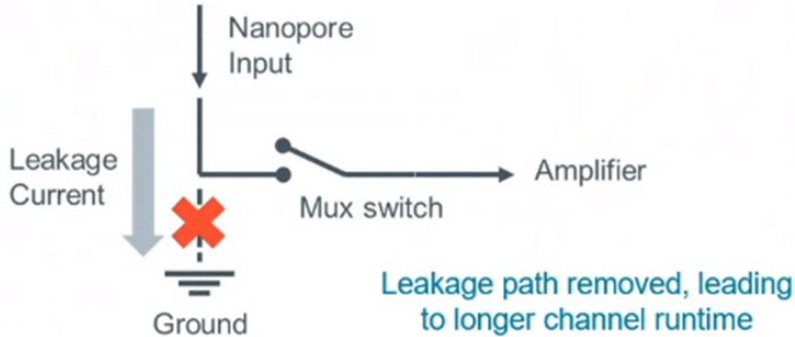
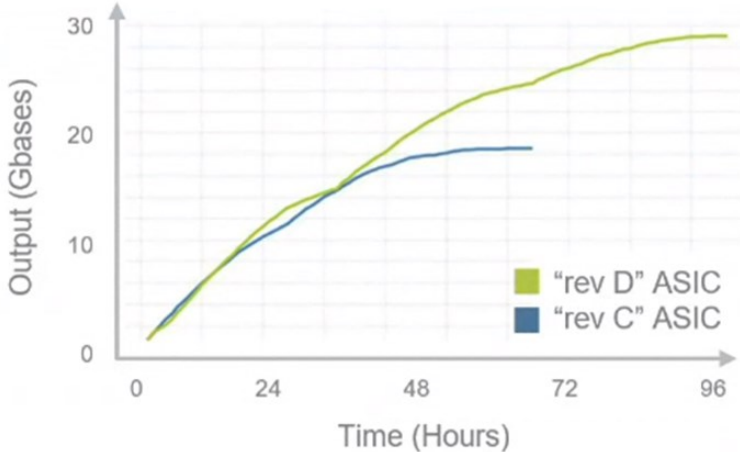
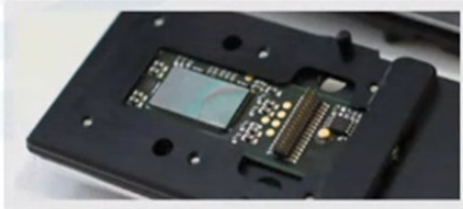
### Enabling longer runtimes: Available Q3/4

Previous ASIC version:

- Small current leakage path in “switched off” wells
- Affected wells that were not being observed
- Chemistry in wells was consumed, run time curtailed
- Approximately 20 % consumption vs active well

Improved ASIC (“rev Ds”)

- Leakage path removed, no losses in other mux groups
- Runtime of flow cell significantly increased



## Remerciements :

### UNRAVELING THE DUPLICATION OF A 35 KB BACTERIAL GENOMIC REGION INVOLVED IN IMMUNOGLOBULIN CLEAVAGE BY LONG READ SEQUENCING

Christophe Boury<sup>1</sup>, Charlotte Mouden<sup>1</sup>, Franck Salin<sup>1</sup>, Erwan Guichoux<sup>1</sup>, Géraldine Gourgues<sup>2,3</sup>, Alain Blanchard<sup>2,3</sup>, Joerg Jores<sup>4</sup>, Carole Lartigue<sup>2,3</sup> and Pascal Sirand-Pugnet<sup>2,3</sup> <sup>1</sup>PGTB, INRA, UMR BIOGECO, Univ. Bordeaux, Cestas, France, <sup>2</sup>INRA, UMR BFP, Villenave d'Ornon, France; <sup>3</sup>Univ. Bordeaux, UMR BFP Villenave d'Ornon, France; <sup>4</sup>Institute of Veterinary Bacteriology, Univ. Bern, Bern, Switzerland

### Impact de la qualité de l'ADN sur le séquençage Oxford Nanopore

Arthur Demené, doctorant de l'UMR Biogeco, sur le champignon pathogène du genre *Castanea* (*Cryphonectriaparasitica*), Sandrine CROS-ARTEIL, INRA, Montpellier

### Human mitochondrial DNA sequencing by Oxford Nanopore MinION

Aurélien Trimouille<sup>1,2</sup>, Charlotte Mouden<sup>3</sup>, Christophe Boury<sup>3</sup>, Armelle Courrèges<sup>4</sup>, Patricia Fergelot<sup>\*1,2,5</sup>, Marie-Laure Martin-Négrier<sup>\*4,6</sup>

<sup>1</sup> CHU Bordeaux, Service de Génétique Médicale, F-33000 Bordeaux, France

<sup>2</sup> Univ. Bordeaux, INSERM 1211, F-33000 Bordeaux, France

<sup>3</sup> INRA, UMR Biodiversity of Genes and Ecosystems, Transcriptomics and Genomics Platform, Cestas, France

<sup>4</sup> CHU Bordeaux, Service de Pathologie, F-33000 Bordeaux, France

<sup>5</sup> Plateforme Génome Transcriptome, Centre de Génomique Fonctionnelle de Bordeaux, Université de Bordeaux, Bordeaux, France.

<sup>6</sup> Univ. Bordeaux, UMR 5293, F-33000 Bordeaux, France; CNRS, Institut des Maladies Neurodégénératives, UMR 5293, F-33000 Bordeaux, France



GÉNOME TRANSCRIPTOME

