



G D E C



Génétique Diversité Ecophysiologie des Céréales

Premiers résultats sur Pacific Bioscience SEQUEL

Véronique GAUTIER



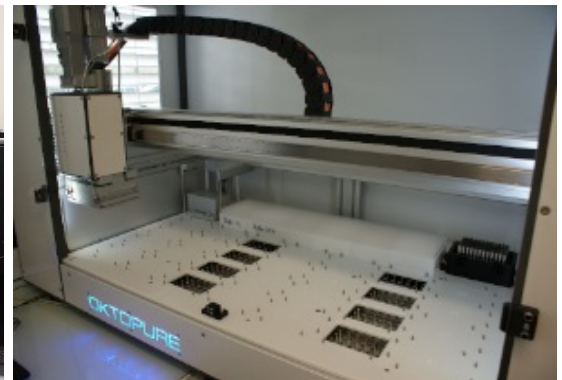
Colloque EPGV Angers 5-7 Octobre 2016

GENTYANE : une équipe, des compétences

- 3 ingénieurs, 2 assistantes ingénieurs, 2 +1 techniciennes (2017)
- Origine des postes : 100% des personnels affecté par le département INRA Biologie et Amélioration des Plantes
- Charles Poncet : manager plateforme
- Anthony Theron et Amélie Bertin : IE chargés de projet (CDD, Prest. Et R/D)
- Elodie Belmonte : AI chargée de projet (CDD)
- Véronique Gautier : AI RMQ et spécialiste NGS
- Lydia Jaffrelo : TR chargée de projets prestation génotypage
- Delphine Boyer : TR chargée de projets développement GDEC
- Chaque agent de l'équipe a en charge le pilotage d'un processus

Une plateforme multiservices

- Génotypage SNP (LC480, Fluidigm Biomark, **Affymetrix Genetitan Axiom automatisée format 96 et 384 exclusivité Gentyane (service provider)**)
- Génotypage Microsatellites (ABI 3730XL)
- Séquençage Massif sur Illumina MiSeq et Pacific Bioscience Sequel (2016)
- Extractions automatisées d'ADN (3200/jour) chimie Sbeadex (plante) et Lifestock (animal) sur robot Oktopure
- Développement de méthodes en collaboration client



Séquenceur Pacific Bioscience Sequel

SEQUEL SYSTEM

Typical Performance

- Average read length: Comparable to PacBio RS II
- Consensus accuracy: Achieves QV50
- Throughput per cell: ~5 – 10 Gb
- SMRT Cells per run: 1 – 16
- Movie lengths: 30 minutes – 6 hours



SMRT SEQUENCING PERFORMANCE

- **Long read lengths**
 - Produce average reads of 10 kb
 - Some reads >60 kb
- **Uniform coverage**
 - No DNA amplification
 - Least GC content and sequence complexity bias
- **High Accuracy**
 - Achieves >99.999% (QV50)
 - Lack of systematic sequencing errors
- **Simultaneous epigenetic detection**
 - Characterize epigenome
 - No separate sample preparation required

Préparation des librairies

- Qualité/Quantité de l'ADN
 - Au moins 10 μ g d'ADN de haut poids moléculaire (proche de 50kb)
 - Sans contaminants (EDTA, détergents, dénaturants, sels, polyphenols...)
 - DO 260/280 entre 1,8 et 2
 - Extraction avec des kits type Macherey ou Qiagen (colonne)

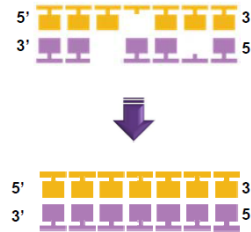
Préparation des bibliothèques

- A partir d'ADN génomique
 - Fragmentation (non nécessaire pour les amplicons)
 - Covaris g-TUBE (> 6Kb)
 - Diagenode Megaruptor

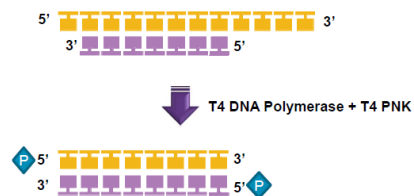


Préparation des librairies

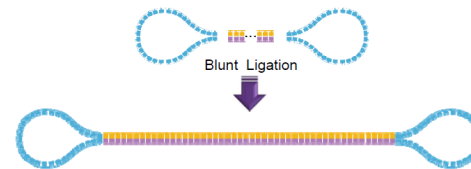
- Réparation de l'ADN



- Réparation des extrémités 5' et 3' et phosphorylation des 5'



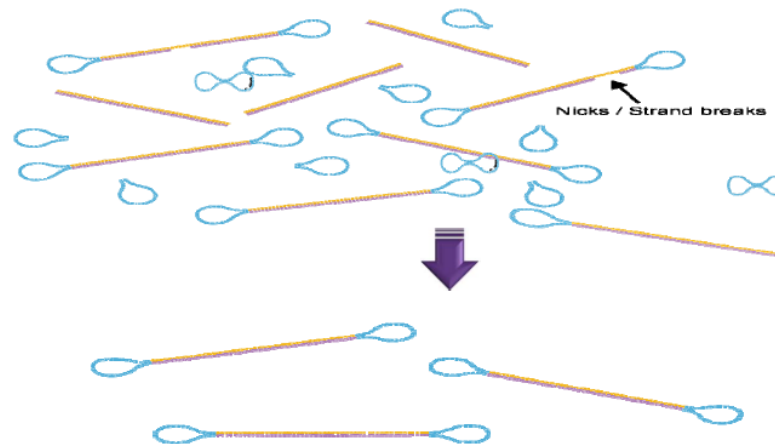
- Fixation des adaptateurs aux 2 extrémités



Préparation des librairies

- Elimination des matrices imparfaites par le traitement à l'EXO III et EXO VII

- Exonuclease III (from 3'-hydroxyl termini)
- Exonuclease VII (from 5'-termini)



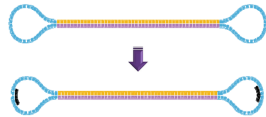
- Library is ready for primer annealing and polymerase binding
- Library can be stored at -20° C for several months

Préparation des librairies

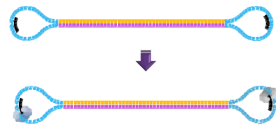
- Sélection de la taille des librairies à l'aide du Sage Science BluePippin

- QC (Qubit+dépôt)

- Fixation des primers de séquençage (attention au ratio)



- Fixation de la polymérase sur les 2 extrémités des SMRTbell template (attention au ratio)



- Fixation des templates sur les MagBeads

- [..\\..\\..\\PacBio\\Content_Pack\\INRA_CEV06_Presentations\\MBS_v5_animation.mov](#)



SMRT Technology

SMRT TECHNOLOGY TEMPLATE FORMAT

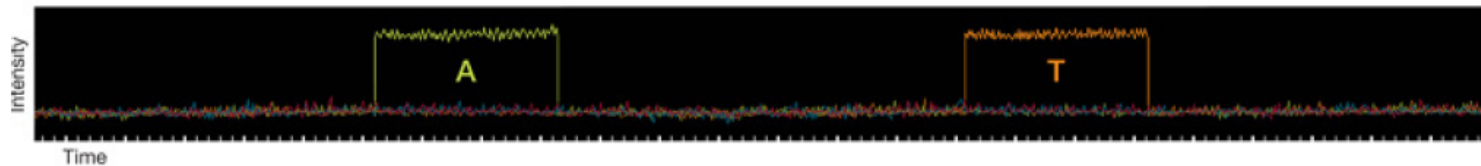


SMRTbell™ Template:

- Structurally linear
- Topologically circular
- Structural homogeneity of templates
- Provides sequences of both forward and reverse strands in the same sequence

SMRT Technology

PROGRESSIVE SYNTHESIS WITH PHOSPHOLINKED NUCLEOTIDES



Step 1: Phospholinked nucleotides are introduced into the zero-mode waveguide (ZMW)

Step 2: The nucleotide is held in the detection volume for tens of milliseconds, fluoresces when excited by light. The captured light is converted into a base call with associated quality metrics

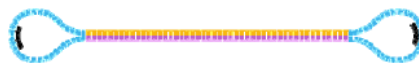
Step 3: The polymerase incorporates the nucleotide, releasing the attached dye molecule

Step 4-5: The process repeats

SMRT Technology

UNIVERSAL SMRTbell™ TEMPLATE

Standard Sequencing for Continuous Long Reads (CLR)



Large Insert Sizes



Generates one pass on each molecule sequenced

Read of Insert / Circular Consensus Sequencing (CCS)



Small Insert Sizes



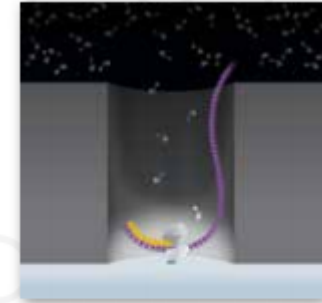
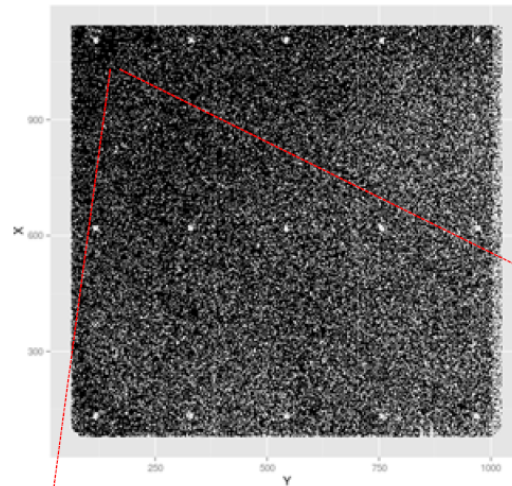
Continued generation of reads per insert size

Generates multiple passes on each molecule sequenced

SMRT Technology



SIGNAL PROCESSING AND BASE CALLING



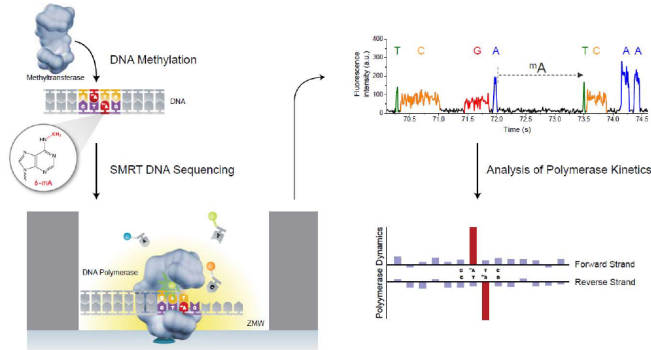
TTACCGGTAGGTACGGGTAAAATGCACCGTTTA

Convert pulses of light in real time into base calls and kinetic measures

Applications

- Séquençage de novo
- Séquençage ciblé :
 - Détection de SNP (possibilité de multiplexer jusqu'à 384 échantillons)
 - Iso-Seq™ Full-Length Transcript Sequencing
- Epigénétique : détection des bases modifiées par la mesure des variations de la cinétique d'incorporation des bases

BASE MODIFICATION: DISCOVER THE EPIGENOME

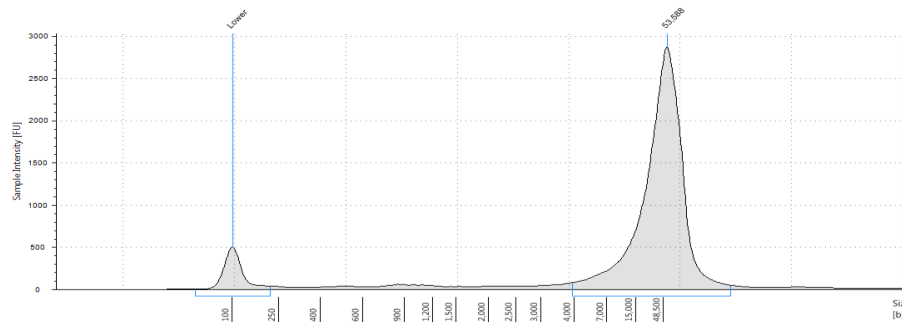


Detect base modifications using the kinetics of the polymerization reaction during sequencing

Premiers résultats

Préparation de librairies à partir d'ADNg de fusarium graminearum (fourni par l'équipe MDC GDEC) et de maïs (CNRGV Toulouse)

- ADN fusarium graminearum (Fg1)
 - Dosage au Qubit : 106 ng/μl
 - Dépôt sur Tape Station Agilent



Préparation des librairies

- Fragmentation à 10 kb
 - Nécessité de partir d'au moins 10 µg d'ADNg pour être sûr d'obtenir 5 µg après fragmentation avec les g-tubes Covaris et purification AMPure

Fusarium (Fg1)	10 µg
Fusarium (FU)	12 µg
Maïs	15 µg

- Traitement exonucléase, réparation, ligation des adaptateurs, purification AMPure, puis sélection sur BluePippin avec un cut-off à 6 Kb

Préparation des librairies

- QC TapeStation et dosage Qubit des différentes étapes

Etape	Fusarium Fg1		Fusarium FU		Maïs	
	Taille	Qtité μg	Taille	Qtité μg	Taille	Qtité μg
Fragmentation	10,6 Kb	10,4	12,3 Kb	13	13,2 Kb	18,6
Avant BluePippin	10,7 Kb	3,9	10,9 kb	4,8	11,7 kb	5,5
Après BluePippin	13,2 Kb	1,2	11,6 Kb	1,6	12,5 Kb	0,9
Librairie finale		1,12		1,02		0,78

Lancement du run

- 3 SMRTCell
- Durée du run : 360 minutes/SMRTCell
- Feuille de route : sur le portail SMRTLink

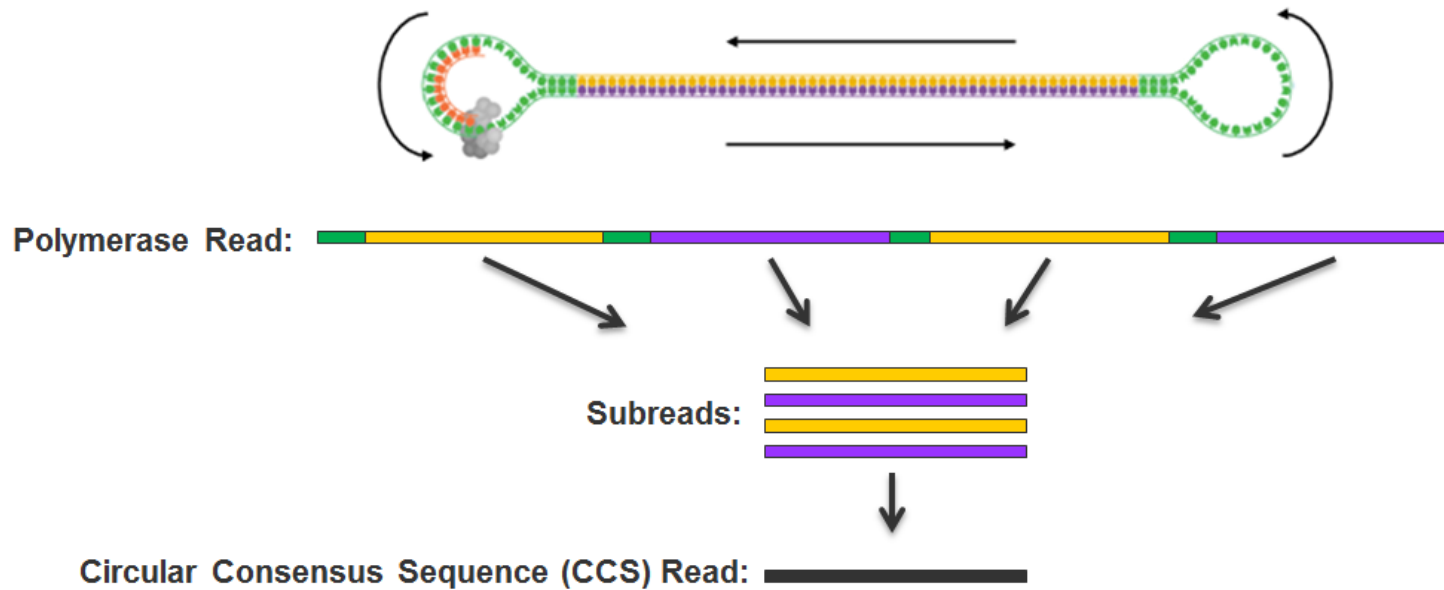
The screenshot displays the PacBio Run Design web interface. The top navigation bar includes the PacBio logo, a menu icon, a help icon, and the title "Run Design". Below the navigation bar, there are buttons for "DELETE", "CANCEL", "VIEW SUMMARY", and "SAVE".

The interface is divided into two main sections:

- Run Information:** This section contains input fields for "Run Name" (with the value "Run 08.18.2016 15:34"), "Run Comments", "Experiment Name", and "Experiment Id". Below these fields is a section titled "Run Reagents / Consumables" which lists the required materials: 3 SMRT Cells, 1 plate Sequencing reagent, 1 tube OS enzyme, 1 tube mineral oil, 3 boxes of tips, 1 mixing plate, and 1 sample plate.
- Sample Information:** This section displays details for three samples. The first sample, "SAMPLE 1: Fusarium G1, A01, 360mins movie, 10000bp insert", is expanded to show its configuration:
 - Sample Name: Fusarium G1
 - Sample Comment: size selected column
 - Sample Well: A 01
 - Mag Bead Loading: ON
 - DNA Control Complex: (empty)
 - Template Prep Kit: SMRTbell™ Template Prep Kit
 - Binding Kit: Sequel™ Binding Kit 1.0
 - Sequencing Kit: Sequel™ Sequencing Plate 1.2
 - Insert Size (bp): 10000
 - Movie Time per SMRT Cell (mins): 360Below this, there is an "Advanced Options" section and two collapsed sample entries: "SAMPLE 2: Fusarium U1, B01, 360mins movie, 10000bp insert" and "SAMPLE 3: Maize cnrgy, C01, 360mins movie, 10000bp insert".

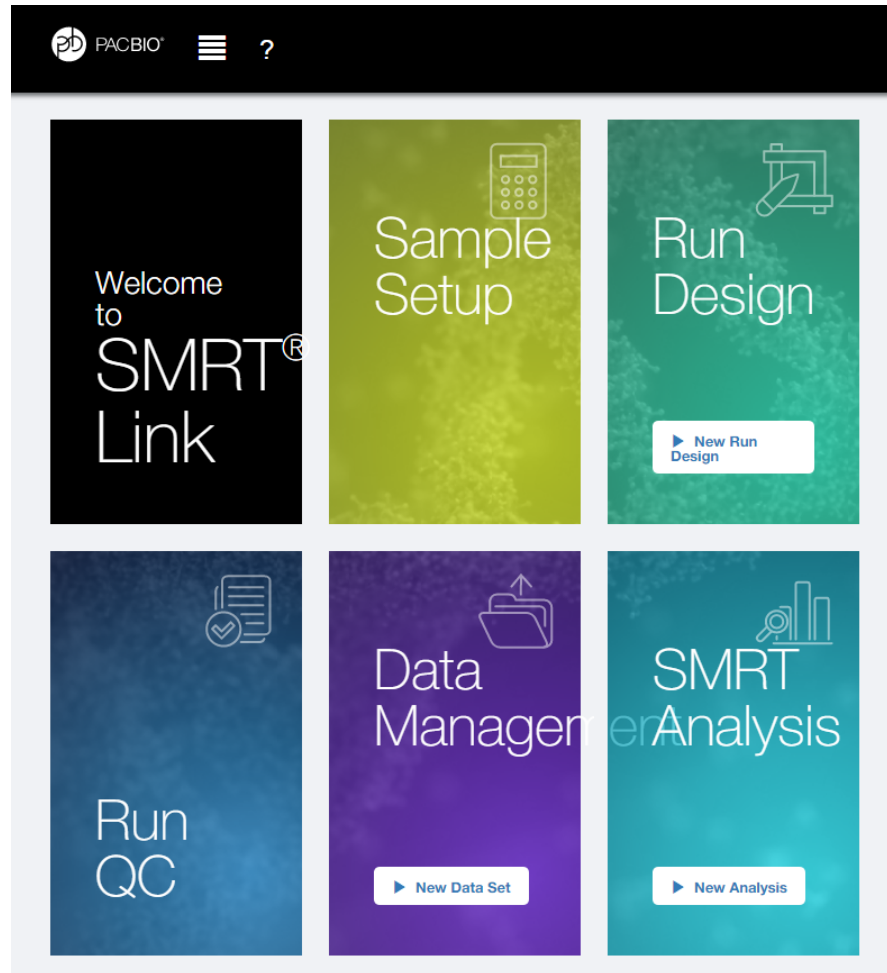
Analyse des runs : Hélène RIMBERT

- Terminologie



Analyse des runs

- A partir de SMRTLink



Analyse des runs

QC

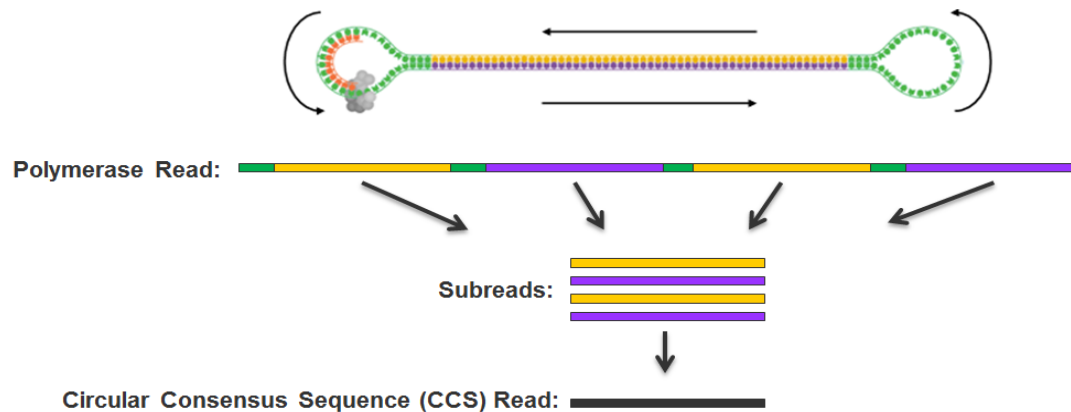
Sample	Fusarium G1	Fusarium U1	Maize CNRGV
Movie time	360	360	360
Total bases (GB)	3,85	3,72	3,41
Polymerase Read Length	7773	7985	6647
Insert Read Length	6186	5874	6032
Productivity P0*	50,60%	53,10%	49,90%
Productivity P1*	47,80%	44,90%	49,50%
Productivity P2*	1,60%	2,00%	0,60%

P0 : The number and percentage of ZMWs that are empty, with no polymerase

P1 : The number and percentage of ZMWs that are productive and sequencing (30<P1 <40%)

P2 : The number and percentage of ZMWs that are not P0 (empty) or P1 (productive). This may occur for a variety of reasons and the sequence data is not usable

Analyse des runs



Analysis Metric	Fg1	Ful	Maize CNRGV
Polymerase Read Bases	3 849 519 705	3 718 187 353	3 409 788 960
Polymerase Reads	495 231	465 671	512 950
Polymerase Read Length (mean)	7 773	7 985	6 647
Polymerase Read N50	12 750	13 250	10 250
Insert Length (mean)	6 186	5 874	6 032

► **Homogénéité des résultats**

Analyse des runs

- Reséquençage de *Fusarium* (Fg1) :
- Séquence de référence : *Fusarium graminearum* PH-1
taille génome : 36 Mb 4 chromosomes + 1 chromosome mitochondrial
- (assembly GCA_000240135.3 http://www.ebi.ac.uk/ena/data/view/GCA_000240135.3)

[Submit this page](#)

Assembly: GCA_000240135.3

ASM24013v3 assembly for *Fusarium graminearum* PH-1

View: [XML](#) [Sequence Report](#)

Download: [XML](#) [Sequence Report](#)

Name	Submitting center	Organism	Strain
ASM24013v3	International <i>Gibberella zeae</i> Genomics Consortium	Fusarium graminearum PH-1	PH-1; NRRL 31084
Assembly level chromosome	Genome representation		
	full		

Description

The [Broad Institute](#) and the International *Gibberella zeae* Genomics Consortium collaborated to sequence the *Gibberella zeae* (aka *Fusarium graminearum*) strain PH-1 genome at approximately 10X coverage using whole genome shotgun (WGS) sequencing. Version 2 of the WGS project consists of 433 nuclear genome contigs in 31 scaffolds, and two mitochondrial contigs/scaffolds. The genome assembly was annotated using automated gene prediction tools, and the assembly was integrated with existing genetic maps.

Lineage

[Eukaryota](#), [Fungi](#), [Dikarya](#), [Ascomycota](#), [Pezizomycotina](#), [Sordariomycetes](#), [Hypocreomycetidae](#), [Hypocreales](#), [Nectriaceae](#), [Fusarium](#)

Navigation

Assembly Statistics

Chromosomes

Assembly Versions

Total length	Ungapped length	Chromosomes & plasmids	Spanned gaps
36,565,771	36,331,366	5	402
Unspanned gaps	Regions	Patches	Alternative loci
12	0	0	0
Scaffolds	Scaffold N50	Contigs	Contig N50
33	5,350,016	435	184,591

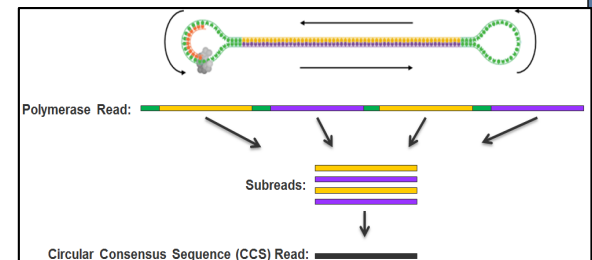
Analyse des runs

- Resequencing for Fusarium G1 : <http://147.99.145.206:9090/#/analysis/job/13>

Analysis Metric	Value
Mean Mapped Concordance	0,83
Number of Subreads (mapped)	720493
Number of Subread Bases (mapped)	3046929877
Subread Length Mean (mapped)	4229
Subread Length N50 (mapped)	6159
Subread Length 95% (mapped)	8590
Subread Length Max (mapped)	31556
Number of Polymerase Reads (mapped)	441271
Polymerase Read Length Mean (mapped)	7064
Polymerase Read N50 (mapped)	12361
Polymerase Read Length 95% (mapped)	23030
Polymerase Read Length Max (mapped)	46949
Mean Coverage	74,37
Missing bases (%)	4,22

► **Quality value**

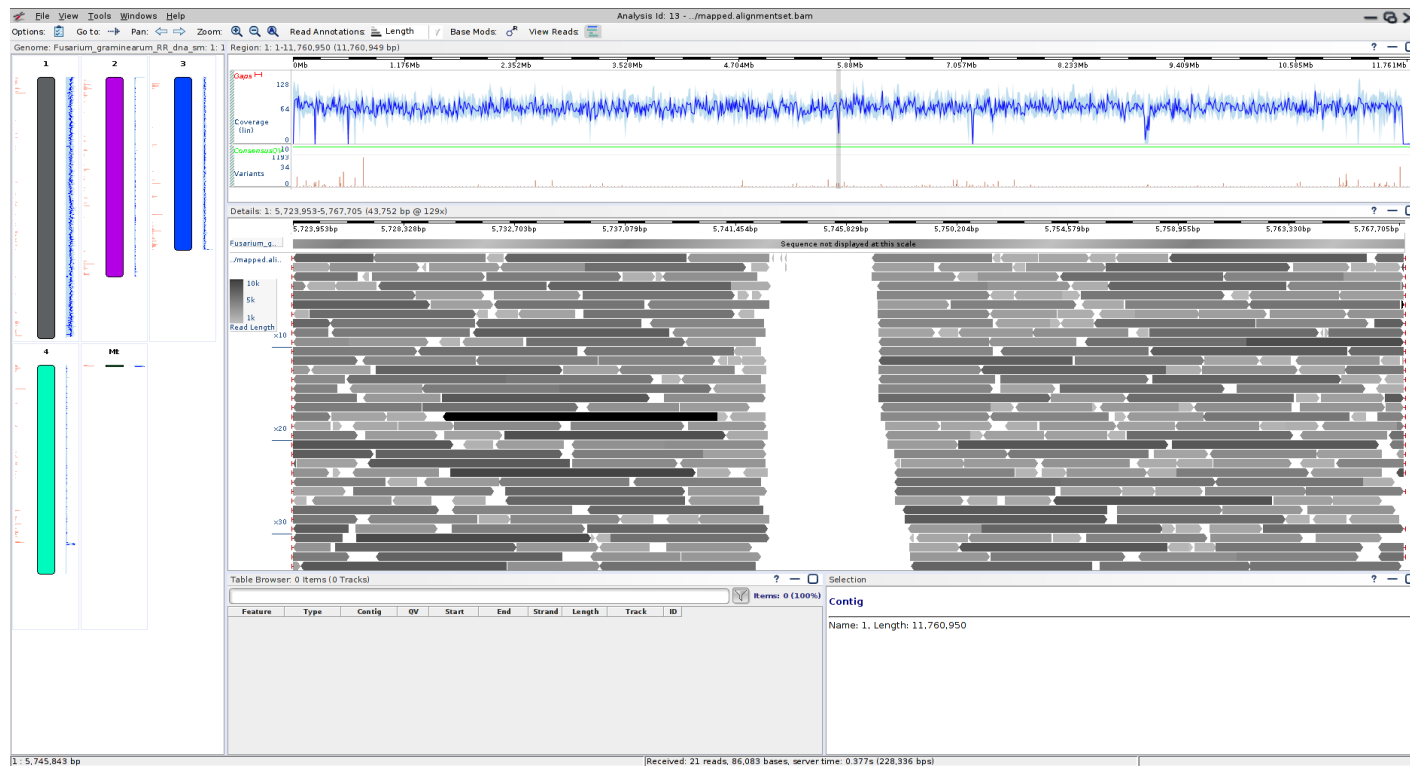
► **soit 100X ≈ de couverture**



Analyse des runs

● Exemple du chromosome 1

- ▶ couverture globalement homogène avec des zones beaucoup moins bien couvertes
- ▶ Mise en évidence d'une zone de chromosome non identique entre la référence et la souche Fg1 séquencée



Analyse des runs

- De novo Assembly HGAP4 Fusarium (FU) :

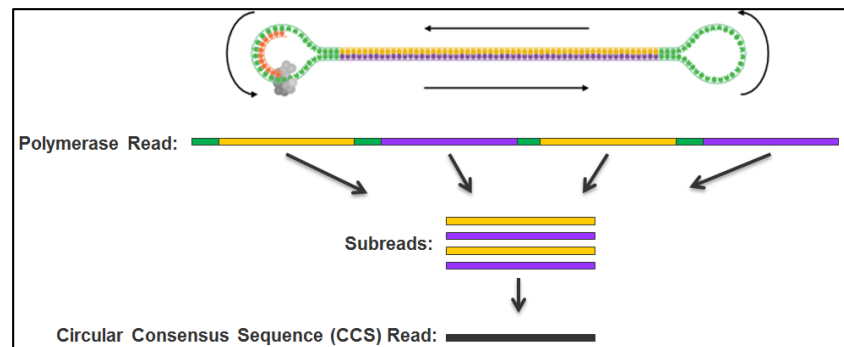
Single-pass
long reads



Select longest as
seed reads



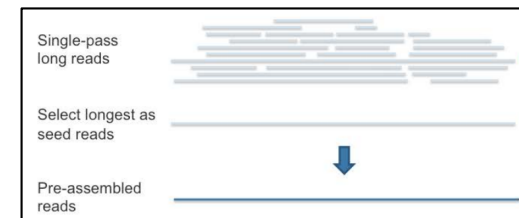
Pre-assembled
reads



De novo Assembly HGAP4 Fusarium (FU) :

- Pre-assembly results

Analysis Metric	Value	
Genome Length (user input)	40 000 000	Subreads
Number of Raw Reads	850 430	
Raw Read Length Mean	4 349,57	
Raw Read Length (N50)	6 136	
Raw Read Length 95%	8 861	
Number of Raw Bases (total)	3 699 007 729	
Raw Coverage (bases/genome_size)	92,48	
Length Cutoff (user input or auto-calc)	6 385	Seed reads :longest single-pass long reads representing 30x coverage.
Number of Seed Reads	142 378	
Seed Read Length Mean	8 430,22	
Seed Read Length (N50)	8 018	
Seed Read Length 95%	13 094	For our dataset, seed length = 6385bp
Number of Seed Bases (total)	1 200 278 397	
Seed Coverage (bases/genome_size)	30,01	1,2Gb → 30x

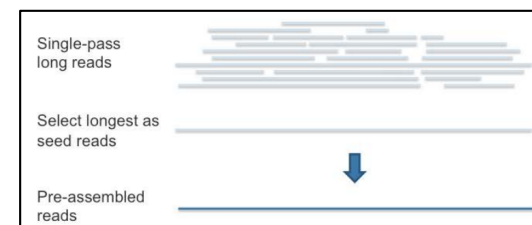


De novo Assembly HGAP4 Fusarium (FU) :

Analysis Metric	Value
Number of Pre-Assembled Reads	134 030
Pre-Assembled Read Length Mean	3 940,05
Pre-Assembled Read Length (N50)	5 286
Pre-Assembled Read Length 95%	7 715
Number of Pre-Assembled Bases (total)	528 084 331
Pre-Assembled Coverage (bases/genome_size)	13,20
Pre-Assembled Yield (bases/seed_bases)	44,00%

Pre-assembled reads :
obtained by mapping
single-pass reads to
seed reads

94 % of seed reads
pre-assembled



De novo Assembly HGAP4 Fusarium (FU) :

Résultats PacBio

Analysis Metric	Value
Polished Contigs	310
Maximum Contig Length	721 775
N50 Contig Length	290 560
Sum of Contig Lengths	36 211 402

génomme référence (HiSeq2000)

Analysis Metric	Value
Contigs	435
Maximum Contig Length	
N50 Contig Length	184591
Sum of Contig Lengths	36 331 366

Conclusion

- Pour du séquençage de novo ou du reséquençage:
 - Préparation de la librairie : 3 jours
 - Quantité de ADNg nécessaire : mini 10 µg
 - Possibilité de passer plusieurs SMRTCell à partir d'une librairie
 - Analyse relativement presse-bouton avec SMRTlink
 - Résultats prometteurs à partir d'ADN de bonne qualité mais extrait classiquement pour Fusarium
 - Coût environ 2000€ HT / SMRTCell

Conclusion

- Pour du séquençage de novo ou du reséquençage:
 - Utilisation de la dernière chimie : P6 et protocole pour des fragments de 20kb
 - Chimie en constante évolution, P7 attendue pour fin 2016 : rendement attendu entre 5 et 10 Gb
 - Résultats pour fusarium obtenus à partir des réglages de base des logiciels d'assemblage et de reséquençage.
 - Pas de nettoyage préalable des datas brutes

Perspectives

- Après des tests très encourageants, évolution vers l'obtention de reads de plus grandes longueurs (shearing, évolution de la polymérase)
- Objectifs : 5 à 10 GB de reads (1 Gb sur RS2) de 15-20kb par SMRTCell pour le premier trimestre 2017
- Projets de séquençage du Lupin blanc sur Sequel
- A venir : prochain user meeting PacBio fin novembre à Barcelone



**Merci de
votre
attention**