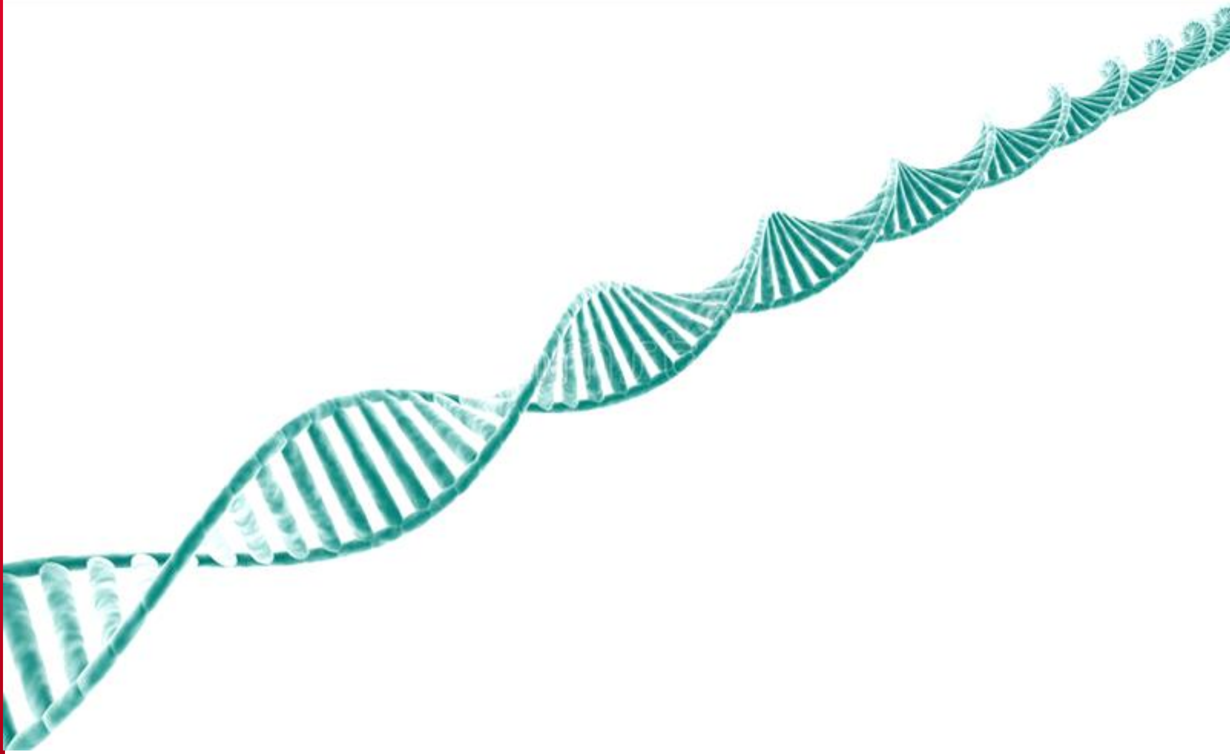
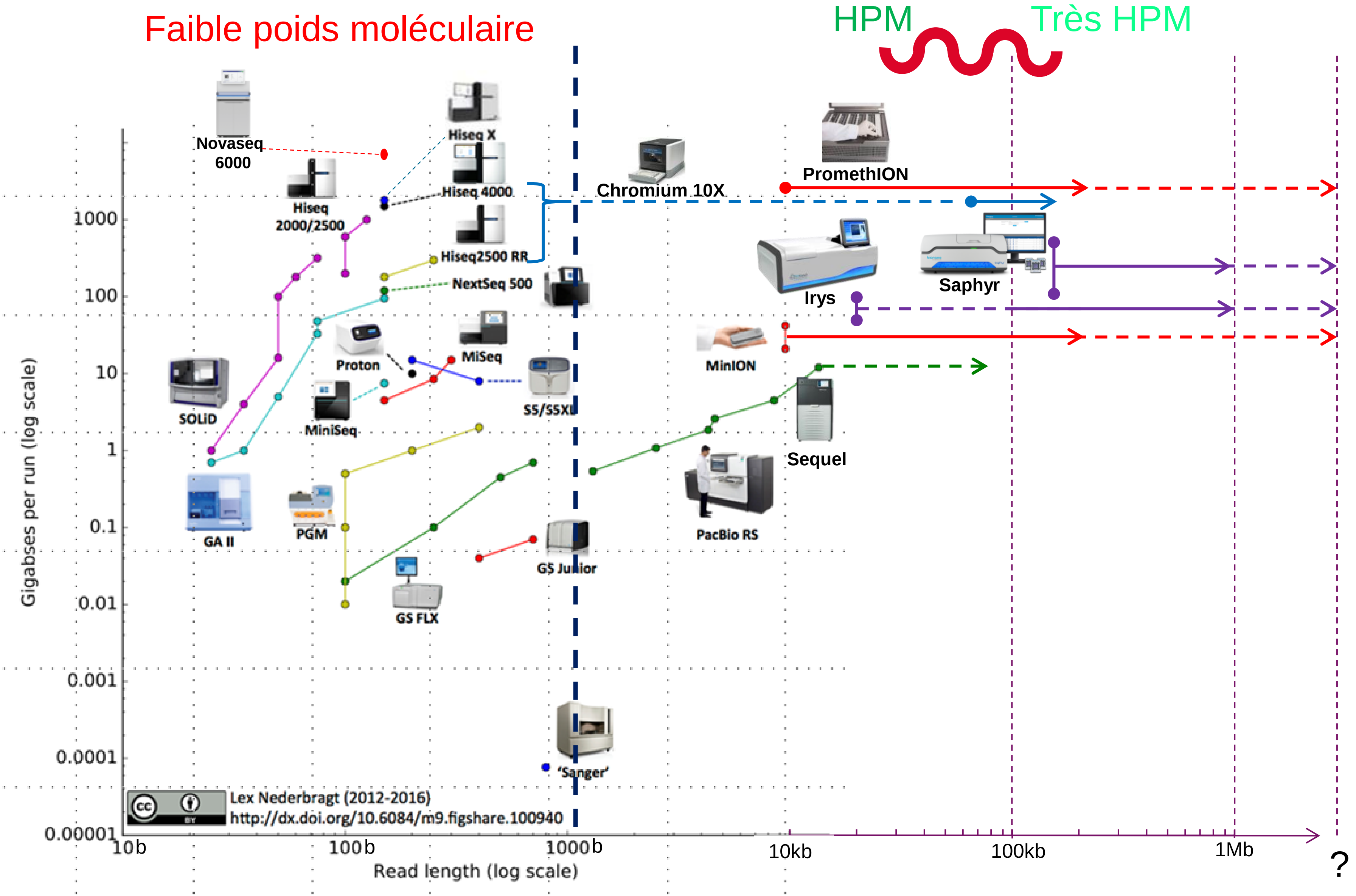


Assemblage *de novo* avec les technologies ONT et Bionano Genomics: l'exemple *Brassica*

Erwan DENIS et Cyril FALENTIN

Colloque EPGV 2018
3,4 et 5 Octobre 2018

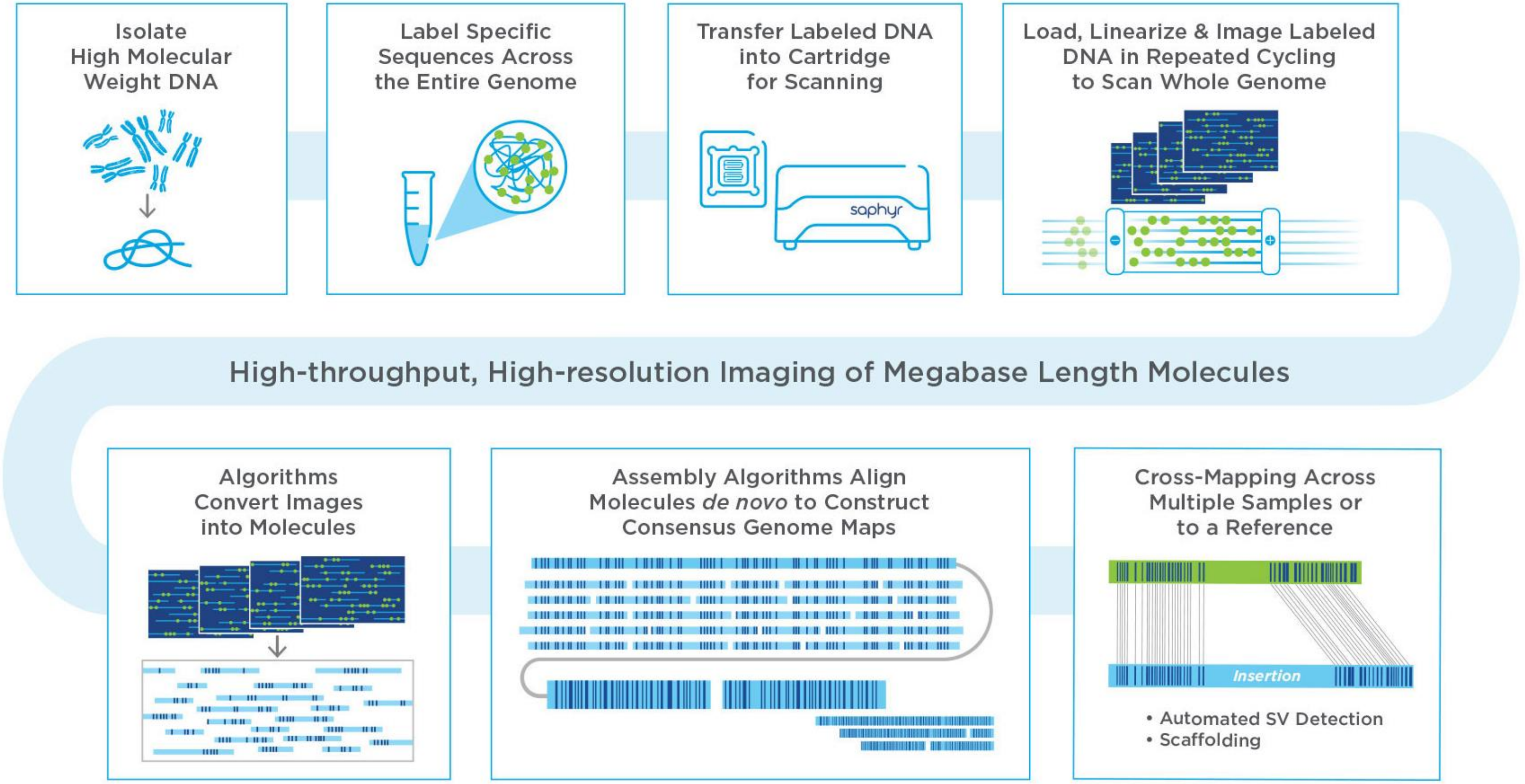
TECHNOLOGIES NECESSITANT DE L'ADN DE HAUT POIDS MOLÉCULAIRE



Lex Nederbragt (2012-2016)
http://dx.doi.org/10.6084/m9.figshare.100940

BIONANO GENOMICS

Technique de cartographie optique



Isolate High Molecular Weight DNA

Label Specific Sequences Across the Entire Genome

Transfer Labeled DNA into Cartridge for Scanning

Load, Linearize & Image Labeled DNA in Repeated Cycling to Scan Whole Genome

High-throughput, High-resolution Imaging of Megabase Length Molecules

Algorithms Convert Images into Molecules

Assembly Algorithms Align Molecules *de novo* to Construct Consensus Genome Maps

Cross-Mapping Across Multiple Samples or to a Reference

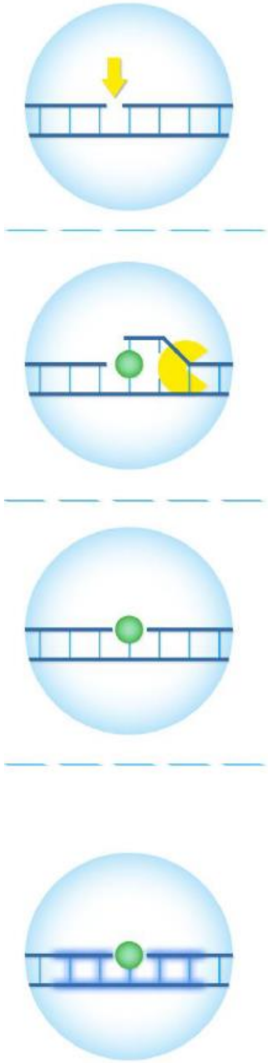
- Automated SV Detection
- Scaffolding

- Avantages: analyse de grandes molécules (150 kb à 2Mb), amélioration des assemblages, détection des SVs, haplotypage, étude de la réplication, étude des marques épigénétiques

- Inconvénients : préparation du matériel difficile, nécessite une référence de qualité.

MARQUAGES BIONANO

NLRS



Nicking

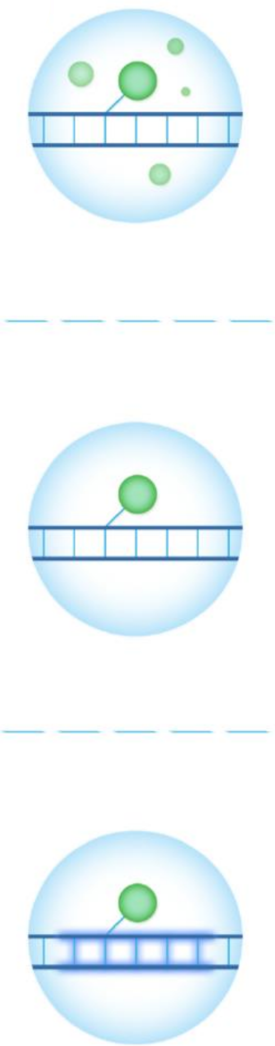
Labeling

Repairing

Staining

DLS

Direct
Labeling
Stain

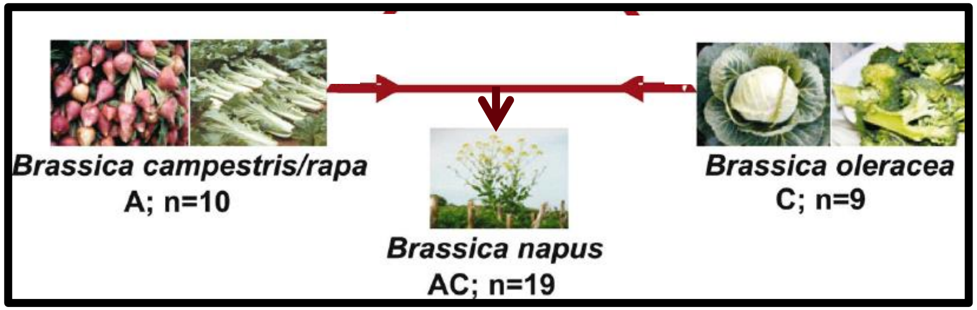


Labeling

Cleanup

Homogenize
& Staining

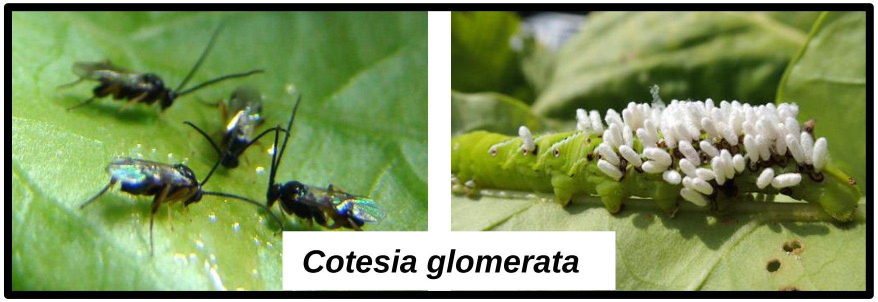
LES DIFFÉRENTS PROJETS



Brassica campestris/rapa
A; n=10

Brassica oleracea
C; n=9

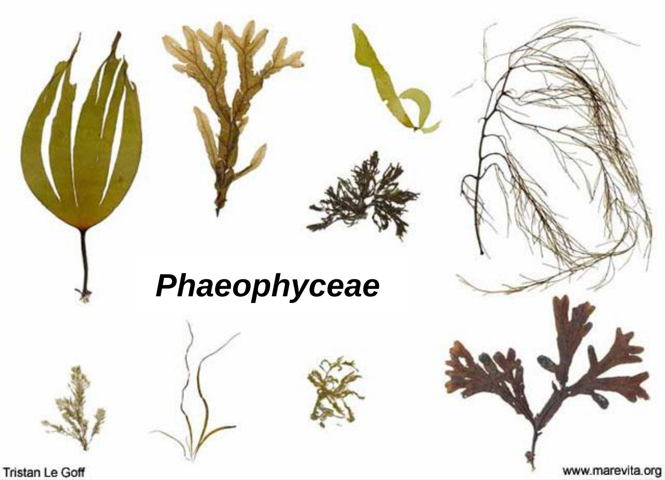
Brassica napus
AC; n=19



Cotesia glomerata

Fagus sylvatica

Populus nigra



Phaeophyceae

Tristan Le Goff

www.marevita.org

Arabidopsis thaliana

Musa spp

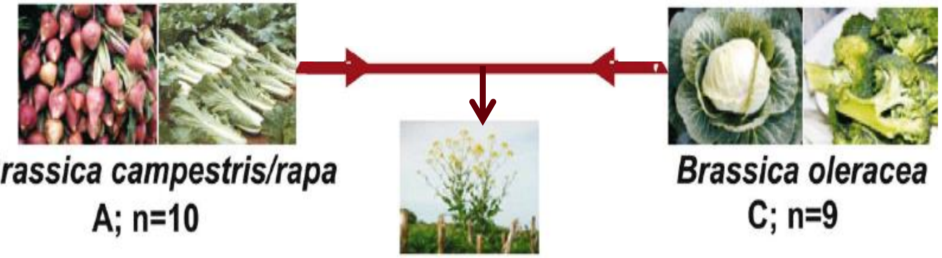
Espeletia schultzii

Citrus spp

Prunus spp

Spartina

THE PROJECT: BIOLOGY



Brassica campestris/rapa
A; n=10
≈ 529Mb

Brassica napus
AC; n=19
≈ 1.2Gb

Brassica oleracea
C; n=9
≈ 630Mb

Human food
and oil production
Oilseed rape 3rd source of
vegetable oil in the world
and 1st in Europe

Musa schizocarpa
S; n= 11

Musa accuminata (A)
Musa balbisiana (B)
AAA, AAB

≈ 600Mb

Human food
>117 millions t/year

Biotic and abiotic stresses: improvement of the selection
Structural variations, phasing: phylogeny, genetic diversity

THE PROJECT: TECHNOLOGIES

Sequencing

Oxford Nanopore Technologies

+

Illumina

MinIon

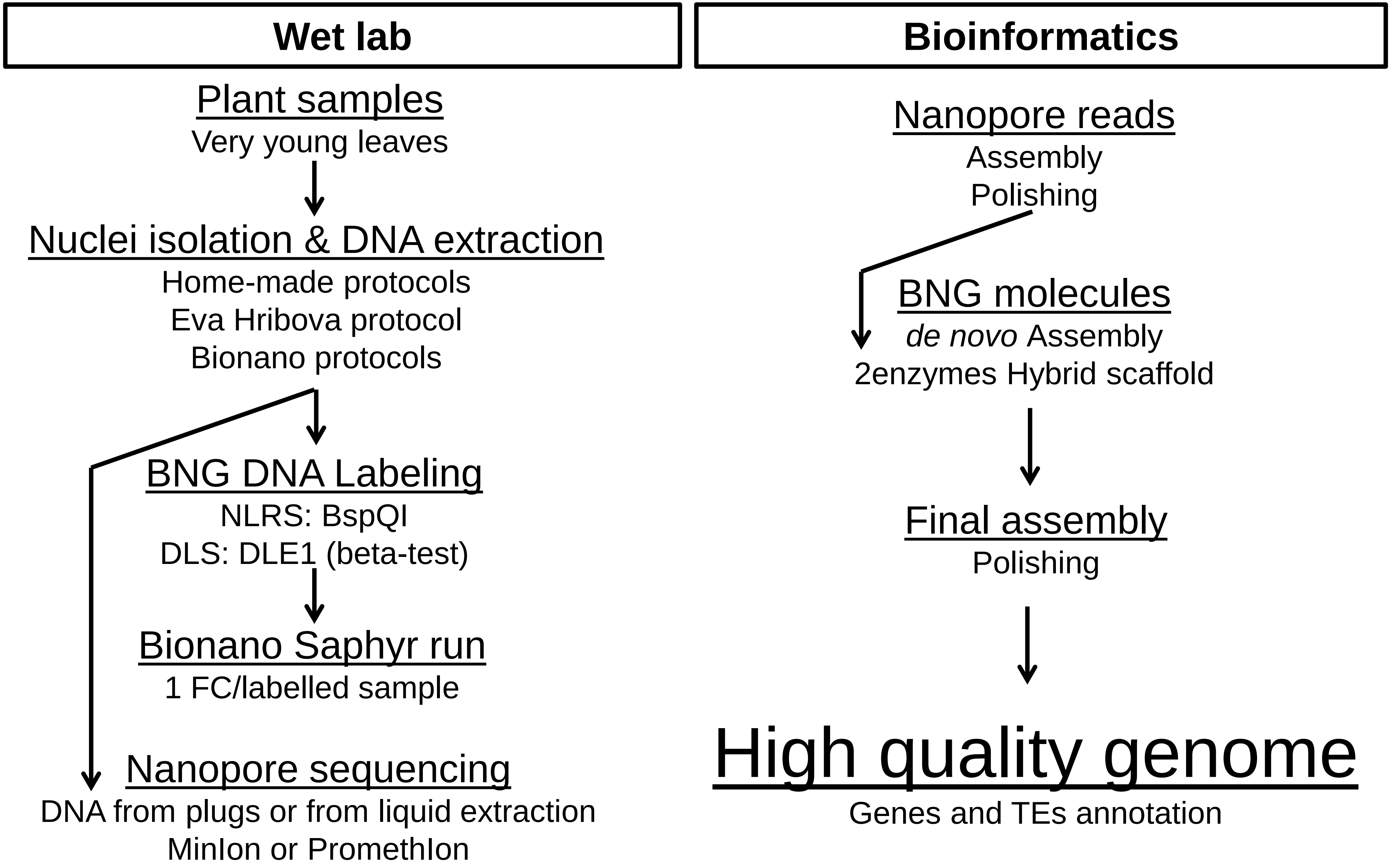
PromethIon

Optical mapping

BioNano Genomics

Saphyr

STRATEGY



Wet lab

Plant samples
Very young leaves

Nuclei isolation & DNA extraction
Home-made protocols
Eva Hribova protocol
Bionano protocols

BNG DNA Labeling
NLRS: BspQI
DLS: DLE1 (beta-test)

Bionano Saphyr run
1 FC/labelled sample

Nanopore sequencing
DNA from plugs or from liquid extraction
MinIon or PromethIon

Bioinformatics

Nanopore reads
Assembly
Polishing

BNG molecules
de novo Assembly
2enzymes Hybrid scaffold

Final assembly
Polishing

High quality genome
Genes and TEs annotation

NUCLEI ISOLATION & DNA EXTRACTION

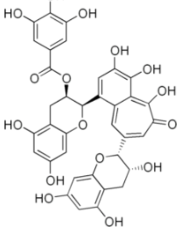
The most limiting step!!!!!

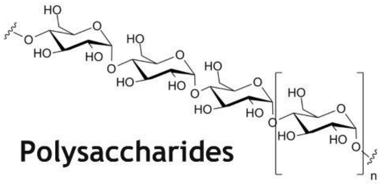

Plants


Young leaves


Dark treatment

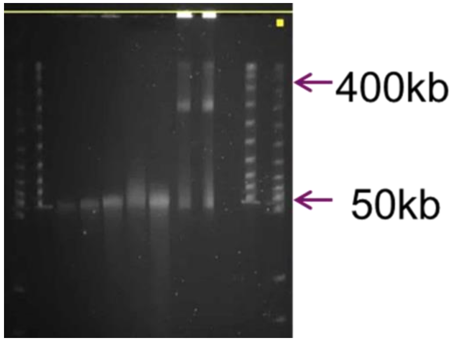
Samples

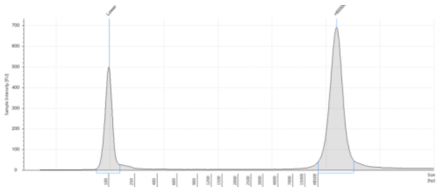

Polyphenols


Polysaccharides

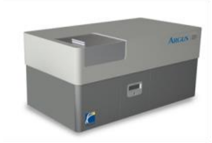
Isolate HMW DNA

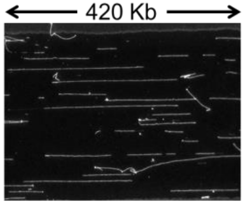
BioNano protocols (4 for plants)
Home-made protocols





Argus® System/Opgen





HMW DNA Quality control

Tapestation (Agilent technologies)
Pippin Pulse (Sage Sciences)
Qcards Argus® System (Opgen)

NUCLEI ISOLATION IMPROVEMENT

For *Brassica spp*:

Home-made protocol adapted from Zhang et al., 1995 and Zhang et al., 2012

Training of the collaborators (INRA, Rennes) at Genoscope

High quality DNA

For *Musa spp*:

Tests on home-made protocols, bionano protocols did not succeed

Test Eva Hribova protocol adaptation: good for *M.schizocarpa*

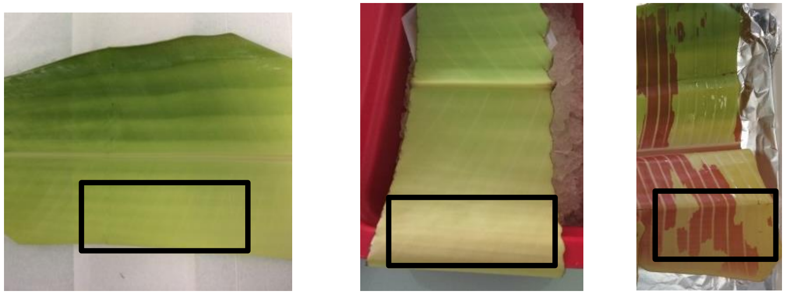
g force	Genome size
1 200xg	> 6Gb
2 000xg	2-6Gb
2 500xg	0.3-2Gb
3 500xg	<0.3Gb

2500g 200g 400g

M. schizocarpa

Good

Other species



Good (but…) Not good Not good

NOUVEAUX ASSEMBLAGES BRASSICA

Brassica rapa et *Brassica oleracea*

+ *Brassica napus* ‘synthétique’ en cours

Brassica rapa : Z1 – yellow sarson – diploïde – 10 chr – ~529Mb
Brassica oleracea : HDEM – broccoli – diploïde – 9 chr – ~630Mb
Brassica napus ‘synthétique’ : RCC-S0 – tétraploïde – 19 chr – ~1,2Gb

Brassica rapa : Z1

Brassica oleracea : HDEM

Brassica napus ‘synthétique’ : RCC-S0

STRATÉGIE MISE EN ŒUVRE POUR ASSEMBLAGE *DE NOVO*

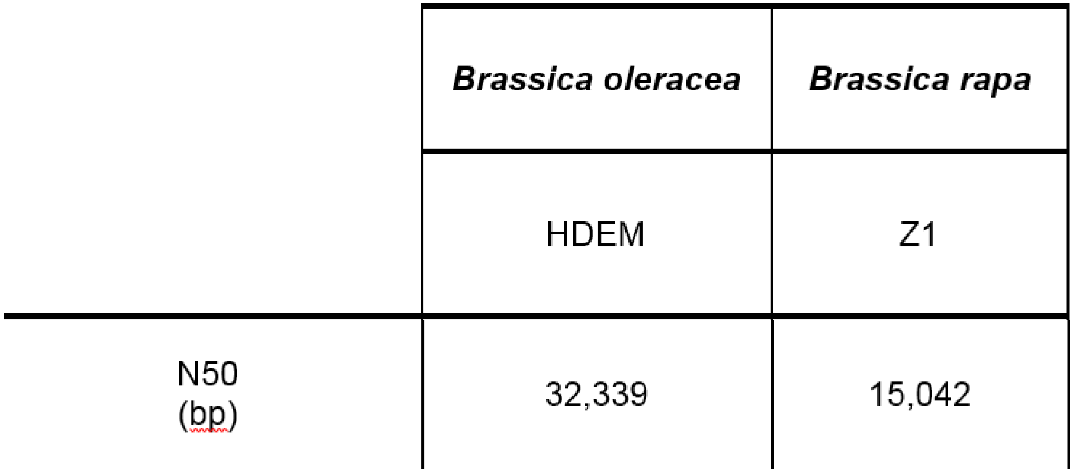
Séquençage et assemblage long-reads Nanopore
entre **32X et 79X**

Constructions de cartes optiques Bionano
2 cartes : ***Bsp*QI et DLE-1**

Séquençage short-reads Illumina : polishing
paired-ends 250pb, PCR-free
environ 50X

STATISTIQUES DES DONNÉES ONT

	Brassica oleracea	*Brassica rapa*
	HDEM	Z1
# Reads (>1Kb)	1,242,230	3,763,611
Cumulative size	20,081,668,921	31,496,053,430
Average size (bp)	16,165.8	8,368.57
N50 (bp)	32,339	15,042
N90 (bp)	7,718	3,570
Max size (bp)	607,633	582,421
# of reads > 50Kb	77,223	30,453
Coverage	31.87X	79X
Coverage (reads >50Kb)	8.23X	5.08X
Number of flowcells	14	12



	Brassica oleracea	*Brassica rapa*
	HDEM	Z1
N50 (bp)	32,339	15,042

STATISTIQUES DES DONNÉES ONT

	Brassica oleracea	***Brassica rapa***
	HDEM	Z1
# Reads (>1Kb)	1,242,230	3,763,611
Cumulative size	20,081,668,921	31,496,053,430
Average size (bp)	16,165.8	8,368.57
N50 (bp)	32,339	15,042
N90 (bp)	7,718	3,570
Max size (bp)	607,633	582,421
# of reads > 50Kb	77,223	30,453
Coverage	31.87X	79X
Coverage (reads >50Kb)	8.23X	5.08X
Number of flowcells	14	12

	Brassica oleracea	***Brassica rapa***
	HDEM	Z1
# of reads > 50Kb	77,223	30,453

STATISTIQUES DES DONNÉES ONT

	Brassica oleracea	*Brassica rapa*
	HDEM	Z1
# Reads (>1Kb)	1,242,230	3,763,611
Cumulative size	20,081,668,921	31,496,053,430
Average size (bp)	16,165.8	8,368.57
N50 (bp)	32,339	15,042
N90 (bp)	7,718	3,570
Max size (bp)	607,633	582,421
# of reads > 50Kb	77,223	30,453
Coverage	31.87X	79X
Coverage (reads >50Kb)	8.23X	5.08X
Number of flowcells	14	12

	Brassica oleracea	***Brassica rapa***
	HDEM	Z1
Number of flowcells	14	12

ASSEMBLAGE *DE NOVO* ONT

11h35

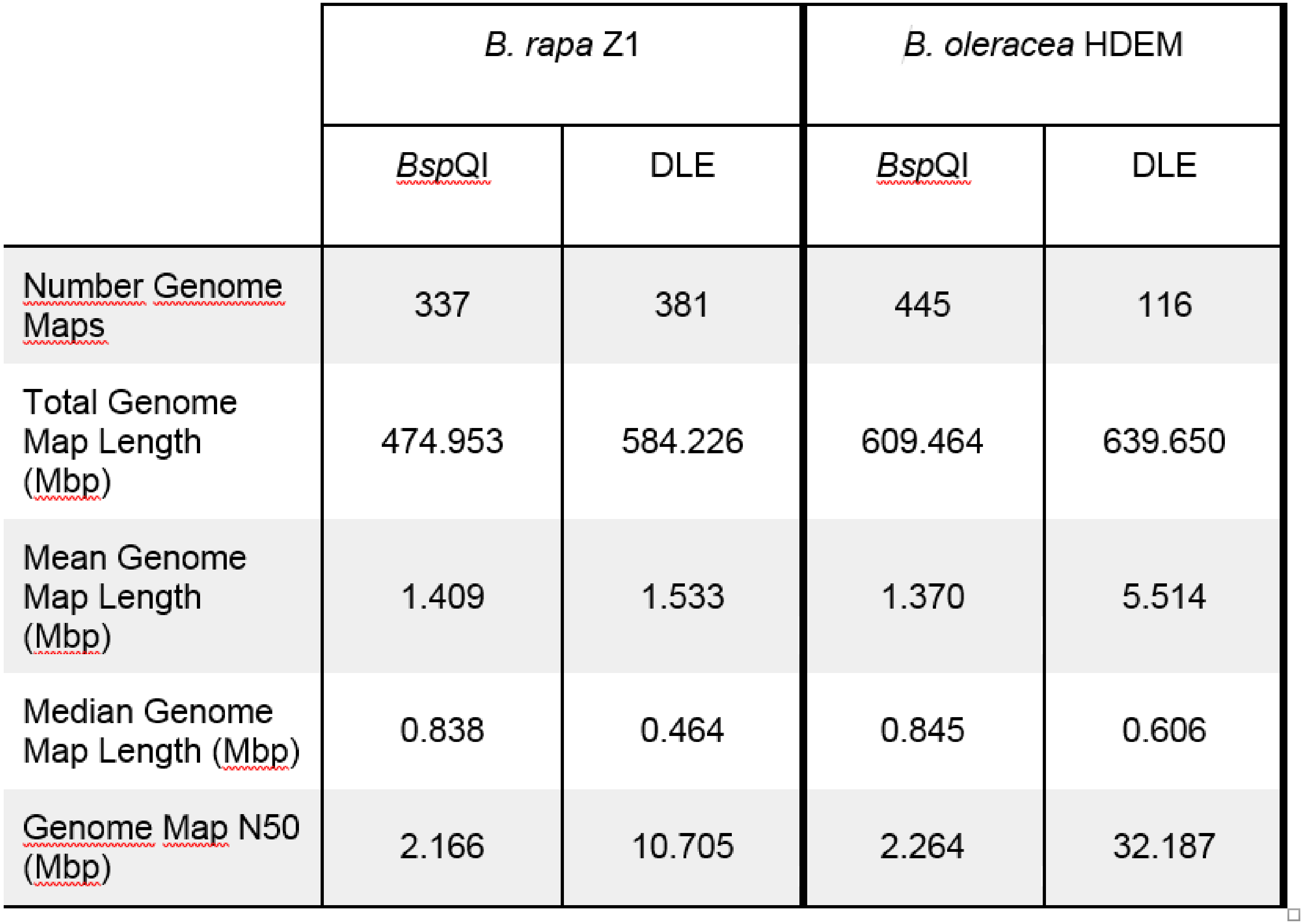
Assemblage de génomes - ONT panorama

Jean-Marc Aury– CEA-IbFJ/Génoscope, Evry

STATISTIQUES DES DONNÉES BRUTES BIONANO

	B. rapa Z1		*B. oleracea* HDEM	
	BspQI	DLE	*BspQI*	DLE
Total number of molecules	3,487,923	12,184,456	6,489,688	9,118,635
Total length (Mbp)	345,728	1,175,790	481,008	480,100
Molecule N50 (kbp)	176	128	177.9	156.3
Label density (/100kb)	9.09	14.29	9.75	13.56

STATISTIQUES DES CARTES OPTIQUES BIONANO

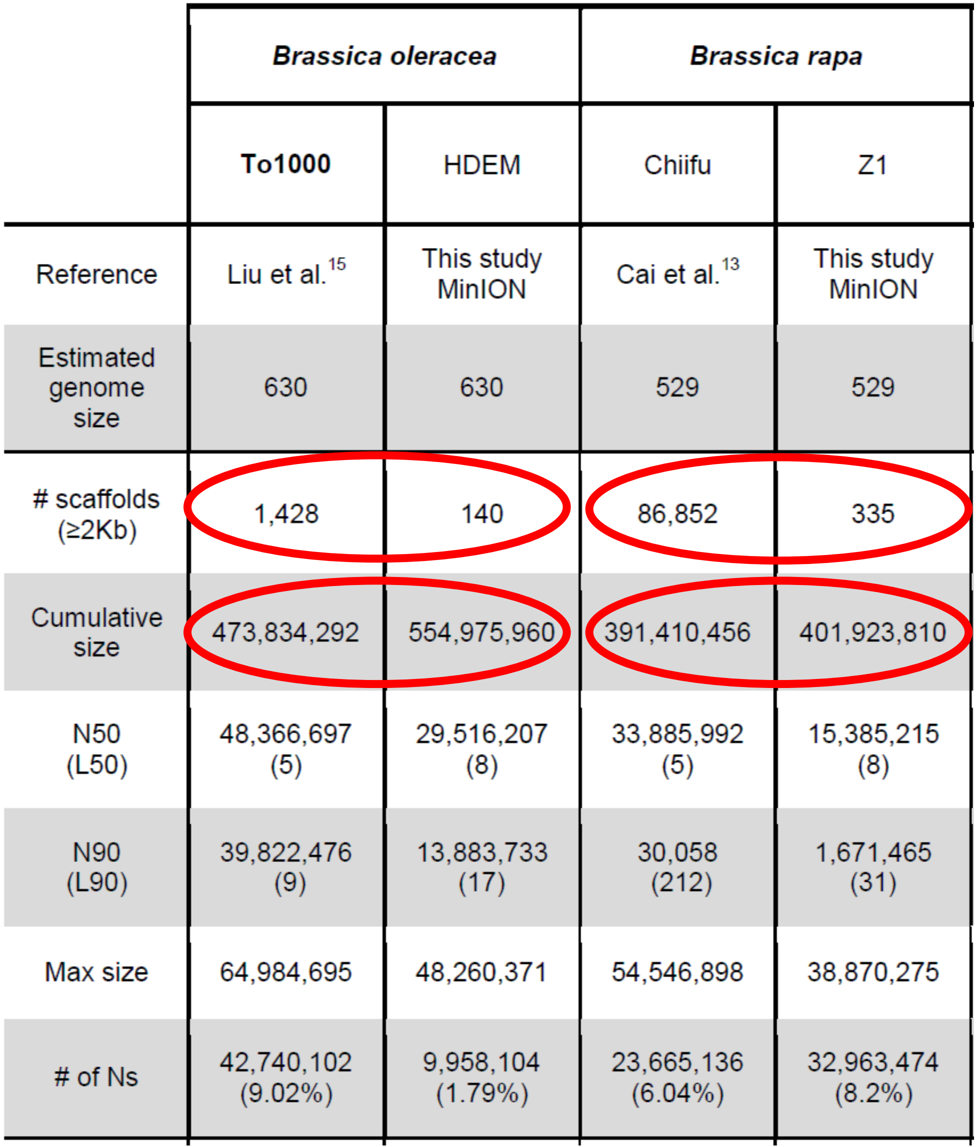


	B. rapa Z1		*B. oleracea* HDEM	
	BspQI	DLE	*BspQI*	DLE
Number Genome Maps	337	381	445	116
Total Genome Map Length (Mbp)	474.953	584.226	609.464	639.650
Mean Genome Map Length (Mbp)	1.409	1.533	1.370	5.514
Median Genome Map Length (Mbp)	0.838	0.464	0.845	0.606
Genome Map N50 (Mbp)	2.166	10.705	2.264	32.187

A rough assembly was 1st performed with parameters: -i 0 -V 0 -A -z -u -m (pipelineCL.py).
This 1st result was used as a reference for a 2nd assembly,
launched with the following parameters: -y -r (rough assembly cmap) -V 0 -m

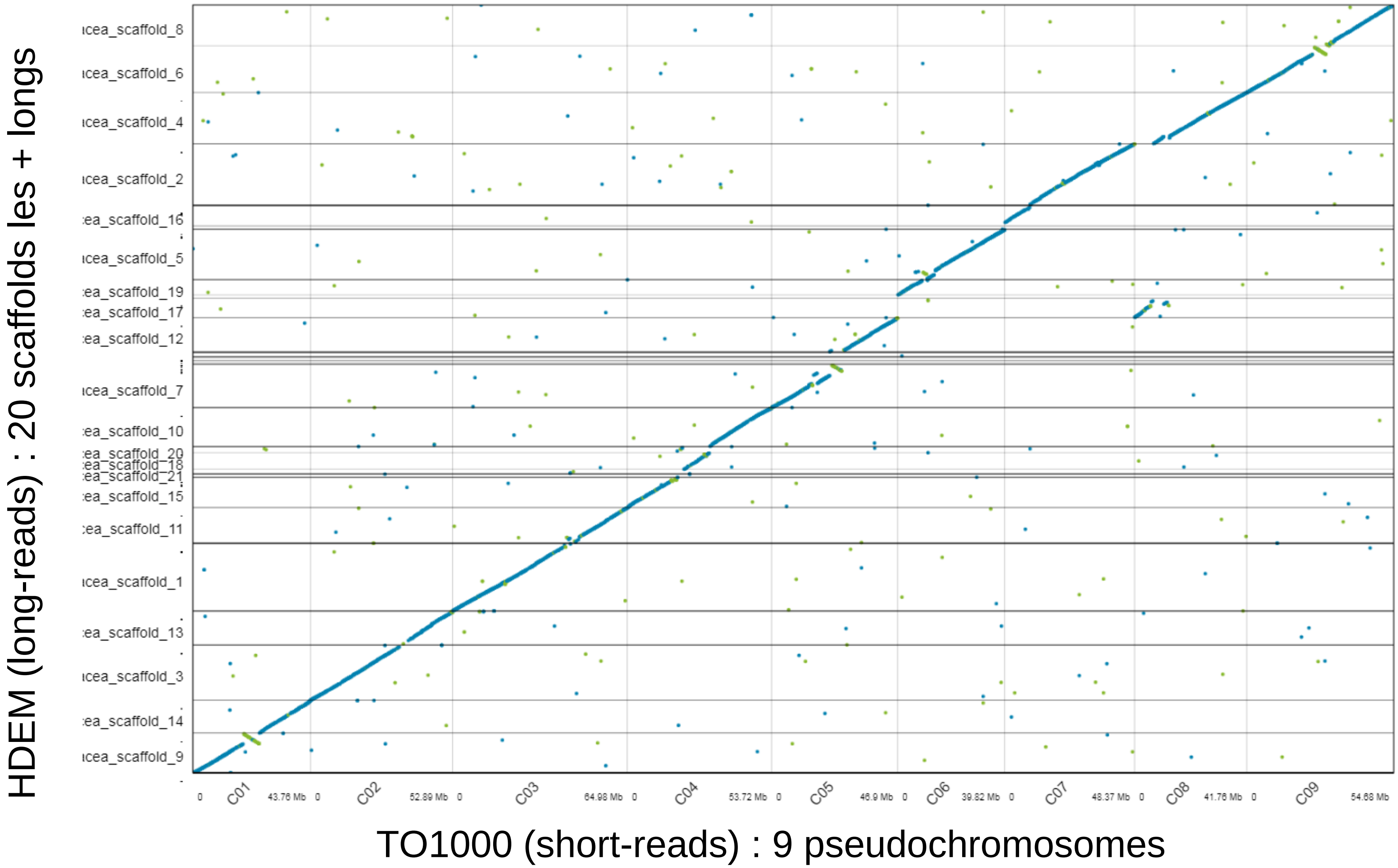
SYNTHÈSE DES SCAFFOLDINGS

Comparaison aux 1er génomes de référence (TO1000 (2014) et Chiifu (2017) : short-reads assemblies)

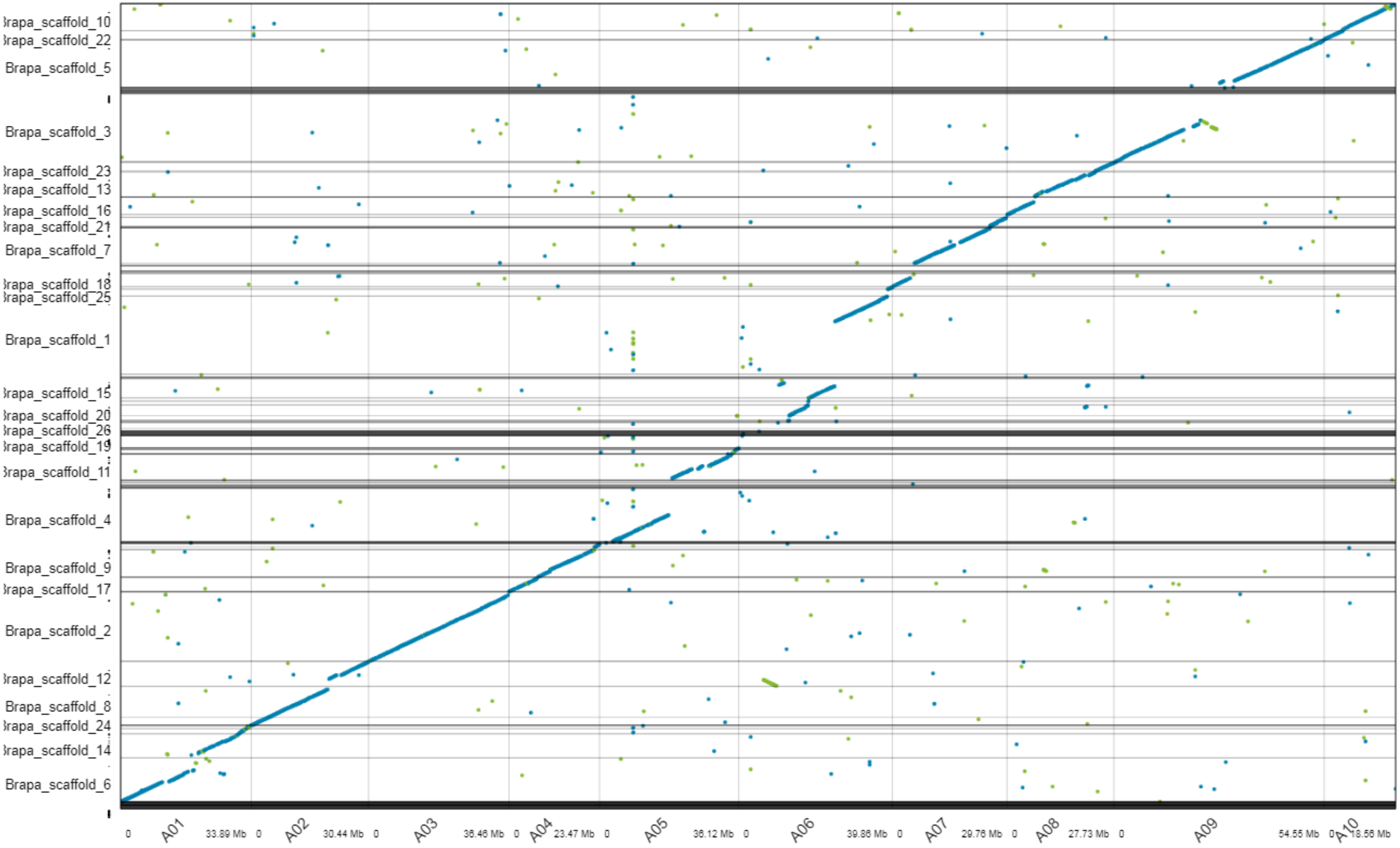


	Brassica oleracea		***Brassica rapa***	
	To1000	HDEM	Chiifu	Z1
Reference	Liu et al.[15]	This study MinION	Cai et al.[13]	This study MinION
Estimated genome size	630	630	529	529
# scaffolds (≥2Kb)	1,428	140	86,852	335
Cumulative size	473,834,292	554,975,960	391,410,456	401,923,810
N50 (L50)	48,366,697 (5)	29,516,207 (8)	33,885,992 (5)	15,385,215 (8)
N90 (L90)	39,822,476 (9)	13,883,733 (17)	30,058 (212)	1,671,465 (31)
Max size	64,984,695	48,260,371	54,546,898	38,870,275
# of Ns	42,740,102 (9.02%)	9,958,104 (1.79%)	23,665,136 (6.04%)	32,963,474 (8.2%)

COMPARAISON GÉNOME TO1000 (SHORT-READS) VS HDEM (LONG-READS) : NUCMER



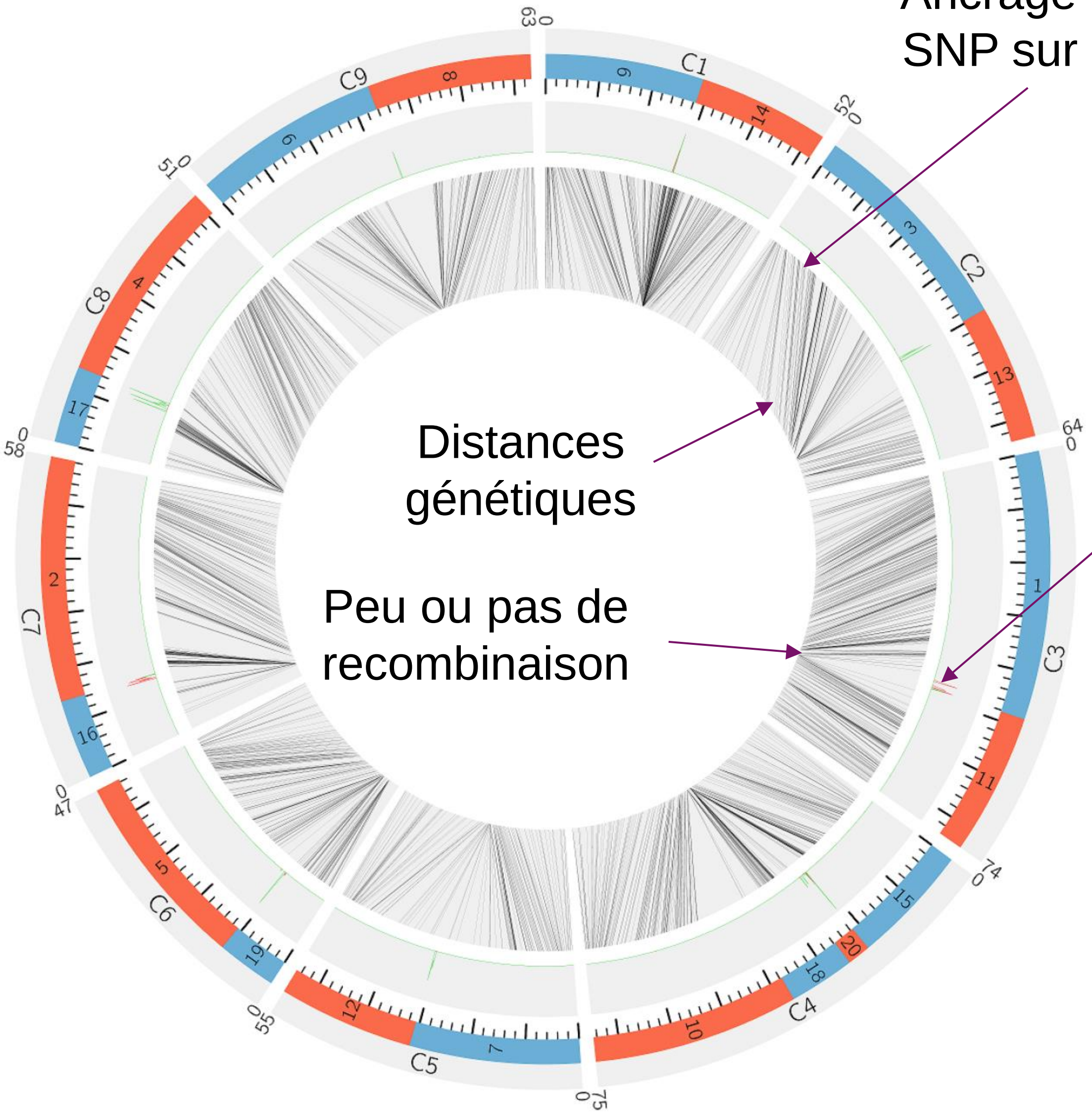
COMPARAISON GÉNOME CHIIFU (SHORT-READS) VS Z1 (LONG-READS) : NUCMER



Z1 (long-reads) : 26 scaffolds les + longs

Chiifu (short-reads) : 10 pseudochromosomes

COMPARAISON ASSEMBLAGE HDEM AVEC LA CARTE GÉNÉTIQUE RICHELAIN X HDEM



Ancrage physique SNP sur scaffolds

Distances génétiques

Ancrage physique de séquences centromériques

Peu ou pas de recombinaison

THANKS

Genoscope: Patrick Wincker

LBioMEG: **Valérie Barbe**
Ghislaine Magdelenat
Barbara Mairey
Wahiba Berrabah

Sequencing Lab: Arnaud Lemainque
Thomas Guerin
Emilie Payen
Corinne Cruaud
Karine Labadie

R&D bioinformatics: Jean-Marc Aury
Caroline Belser
Benjamin Istace
Marion Dubarry
Benjamin Noel
Stefan Engelen
France Denoeud

INRA:
Biodiversity and Polyploidy team (part of Brassica group) : Anne-Marie Chèvre
Gwenaëlle Deniot

CIRAD:
Structure et évolution des génomes (Musa team): D’Hont Angélique
Franc-Christophe Baurens
Guillaume Martin