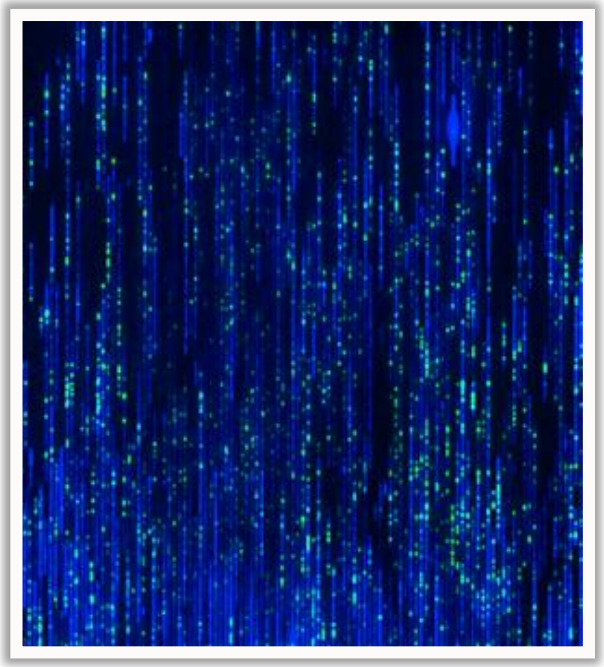
INRA SCIENCE & IMPACT
PLANT GENOMIC CENTER

Optical mapping at the French Plant Genomic Center with the BioNano Irys system

Céline Jeziorski
HeliOr meeting
November 23th, Cordoba



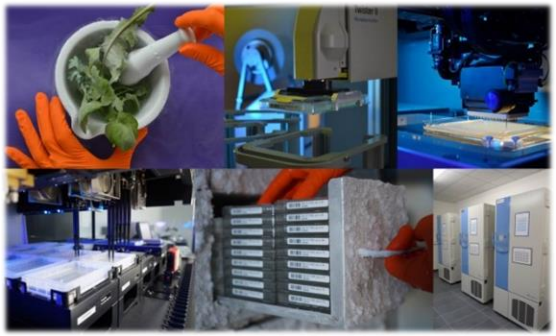
The French Plant Genomic Center

ISO 9001:2008
Octobre 2005

Created in 2004 by INRA (French National Institute for Agricultural Research)

- **Depository of genomic libraries for the scientific community**

⇨ BAC libraries



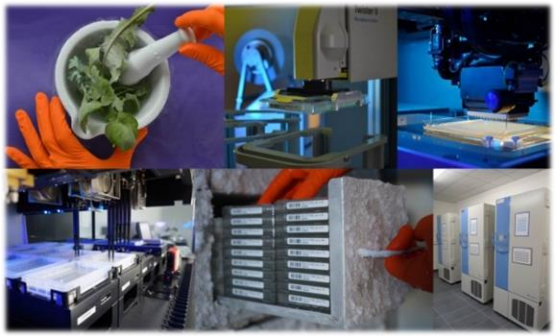
The French Plant Genomic Center

ISO 9001:2008
Octobre 2005

Created in 2004 by INRA (French National Institute for Agricultural Research)

- **Depository of genomic libraries for the scientific community**

⇨ BAC libraries

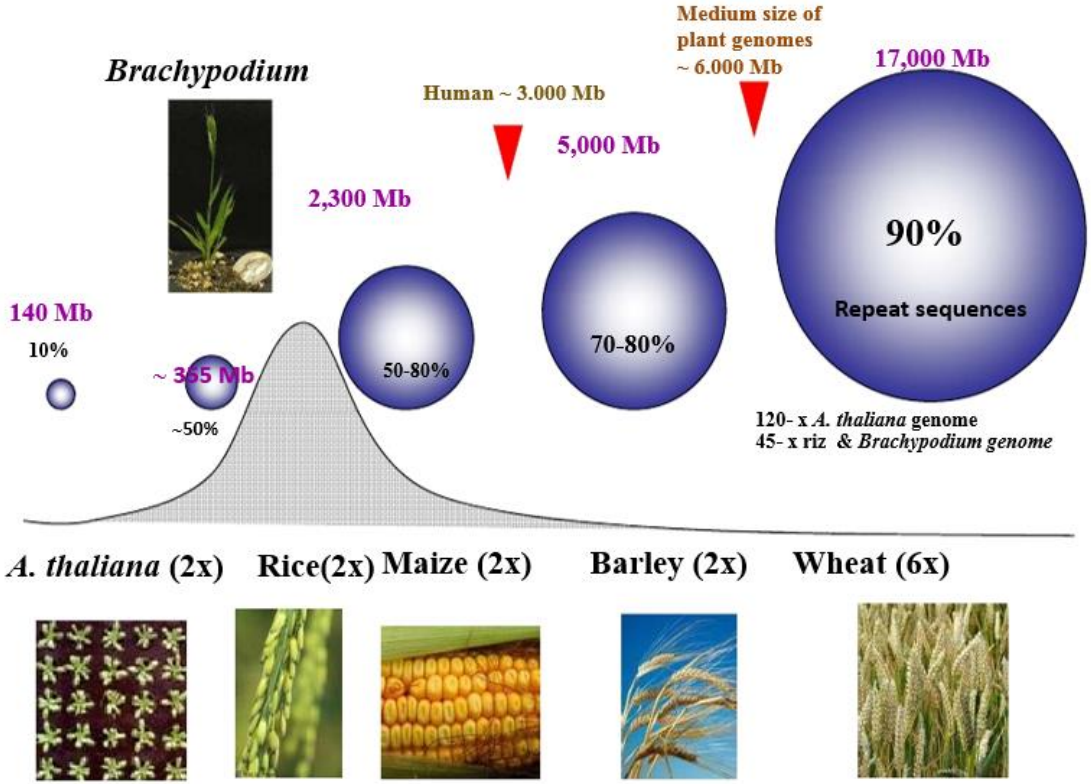


- **A dedicated structure to assist plant genomic programs**
  - ⇨ Distribute the genomic resources at the international level
  - ⇨ Provide high quality research material and efficient tools to study plant genomes
  - ⇨ Develop innovative solutions
  - ⇨ Develop genomic projects in collaboration
  - ⇨ Host scientists in the frame of collaborations

Interactions with laboratories around the world

- **3 233 242 BAC clones distributed during the last 5 years (2011-2015)**

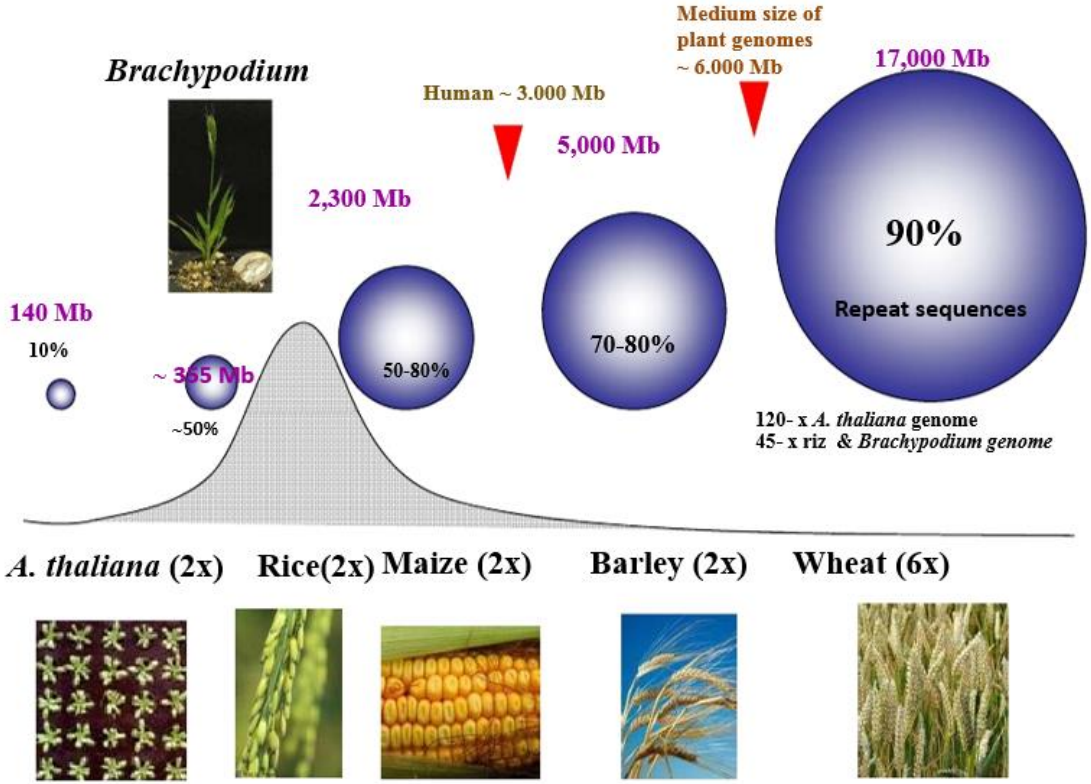
The complexity of plant genomes



Plant genomes are challenging:

- **Genome size**
- **Repeat elements (>80% in maize, barley, sunflower…)**
- **Polyploïdy**

The complexity of plant genomes



Plant genomes are challenging:

- **Genome size**
- **Repeat elements (>80% in maize, barley, sunflower…)**
- **Polyploïdy**

The use of BAC libraries is important

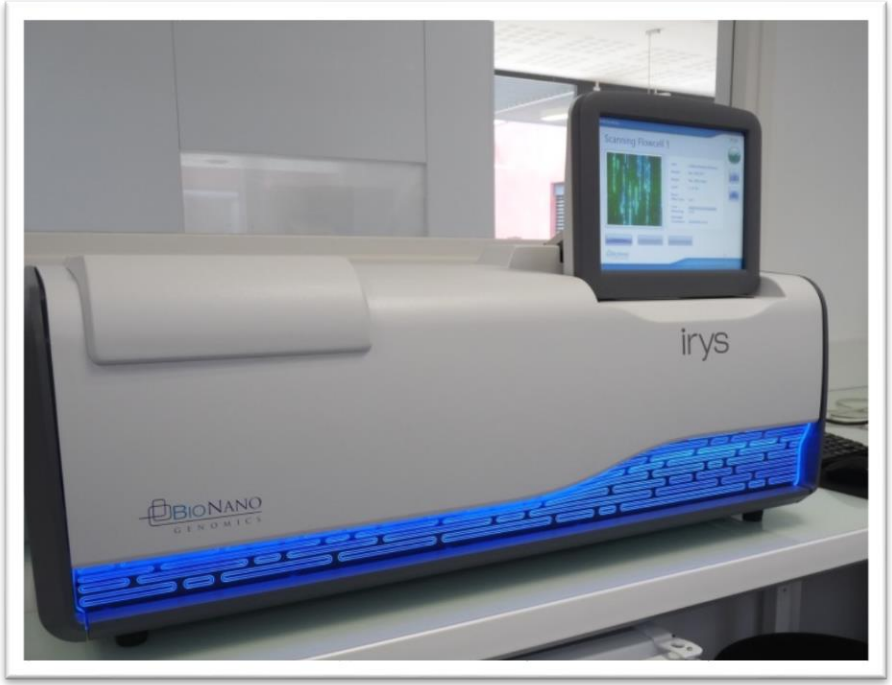
- **Reduce the complexity**
- **Focus on specific target/markers**
- **Manage the diversity**

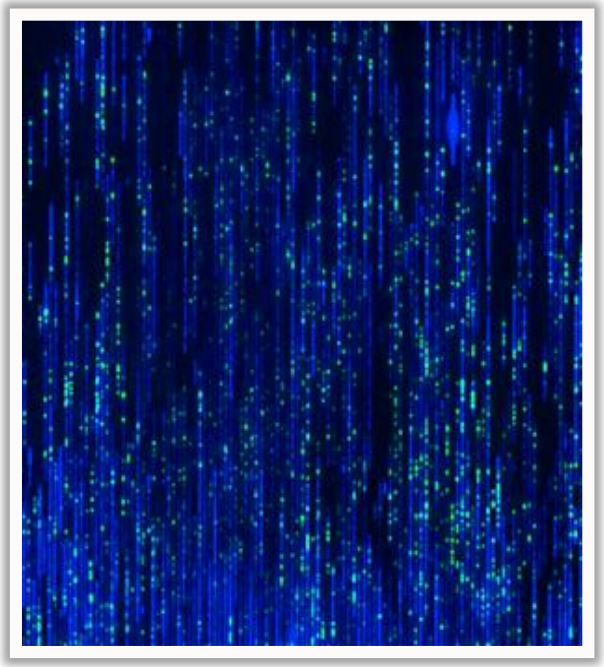
Focus on the optical mapping with BioNano

The Bionano Irys system: another tool to better study the complex genomes

Advantages of BioNano optical mapping:

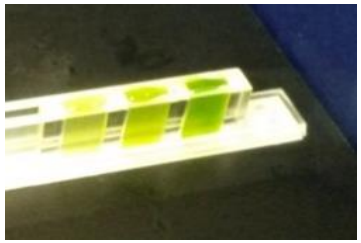
-Direct visualization of long DNA molecules (>100 kb)
-Physical structure and map
-Provides real physical distance information

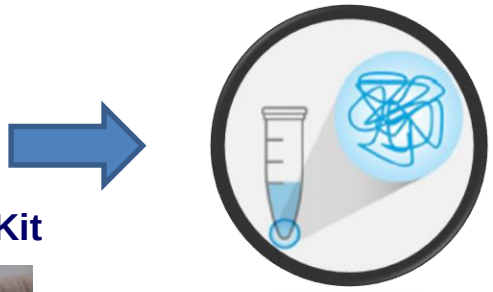


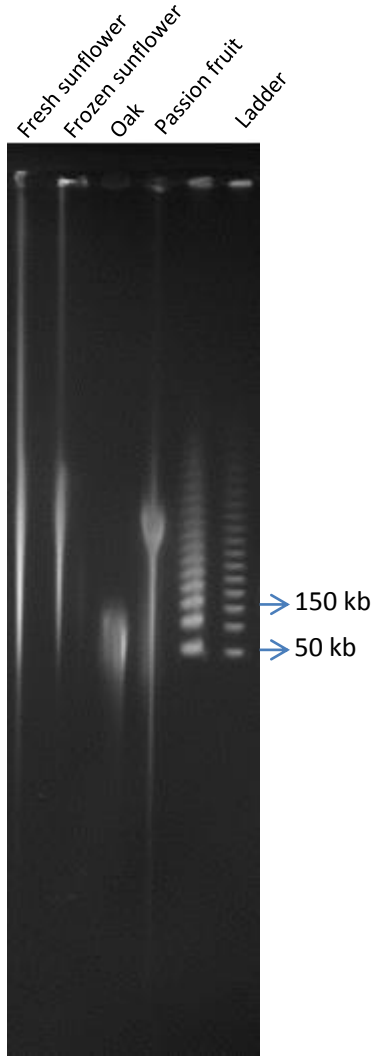


Isolation of High Molecular Weight DNA

Using Bionano Plant DNA Isolation Kit

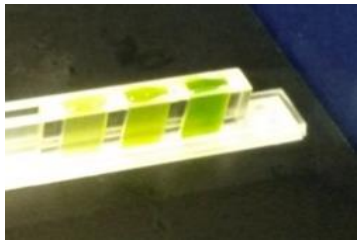


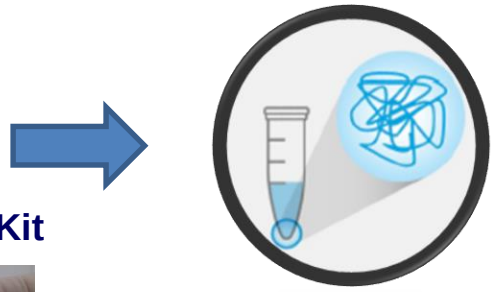


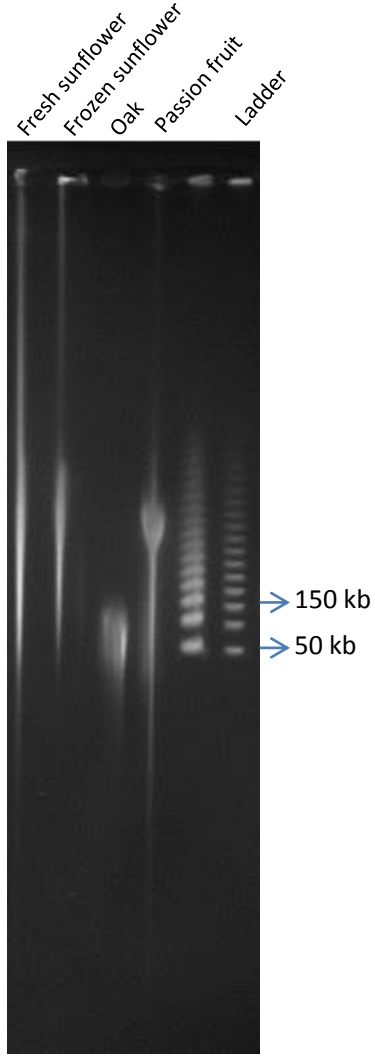


Isolation of High Molecular Weight DNA

Using Bionano Plant DNA Isolation Kit

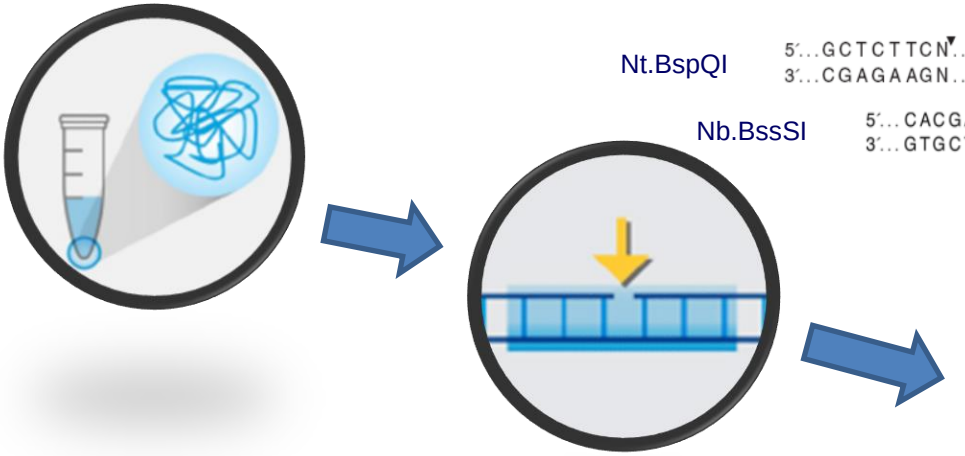






Important to have a high quality of HMW DNA

Nick Label Repair Stain (NLRS) workflow

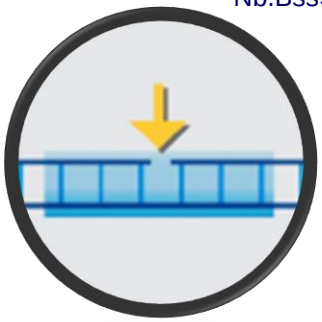


A nicking enzyme creates single strand nicks at recognition sites

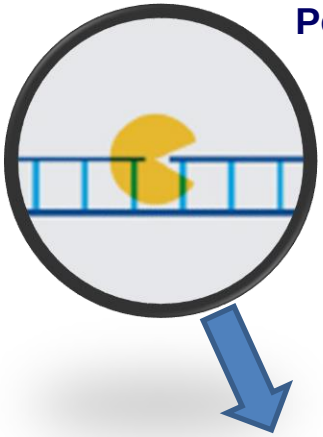
Nick Label Repair Stain (NLRS) workflow

Nt.BspQI 5′…GCTCTTCN…3′ / 3′…CGAGAAGN…5′

Nb.BssSI 5′…CACGAG…3′ / 3′…GTGCTC…5′

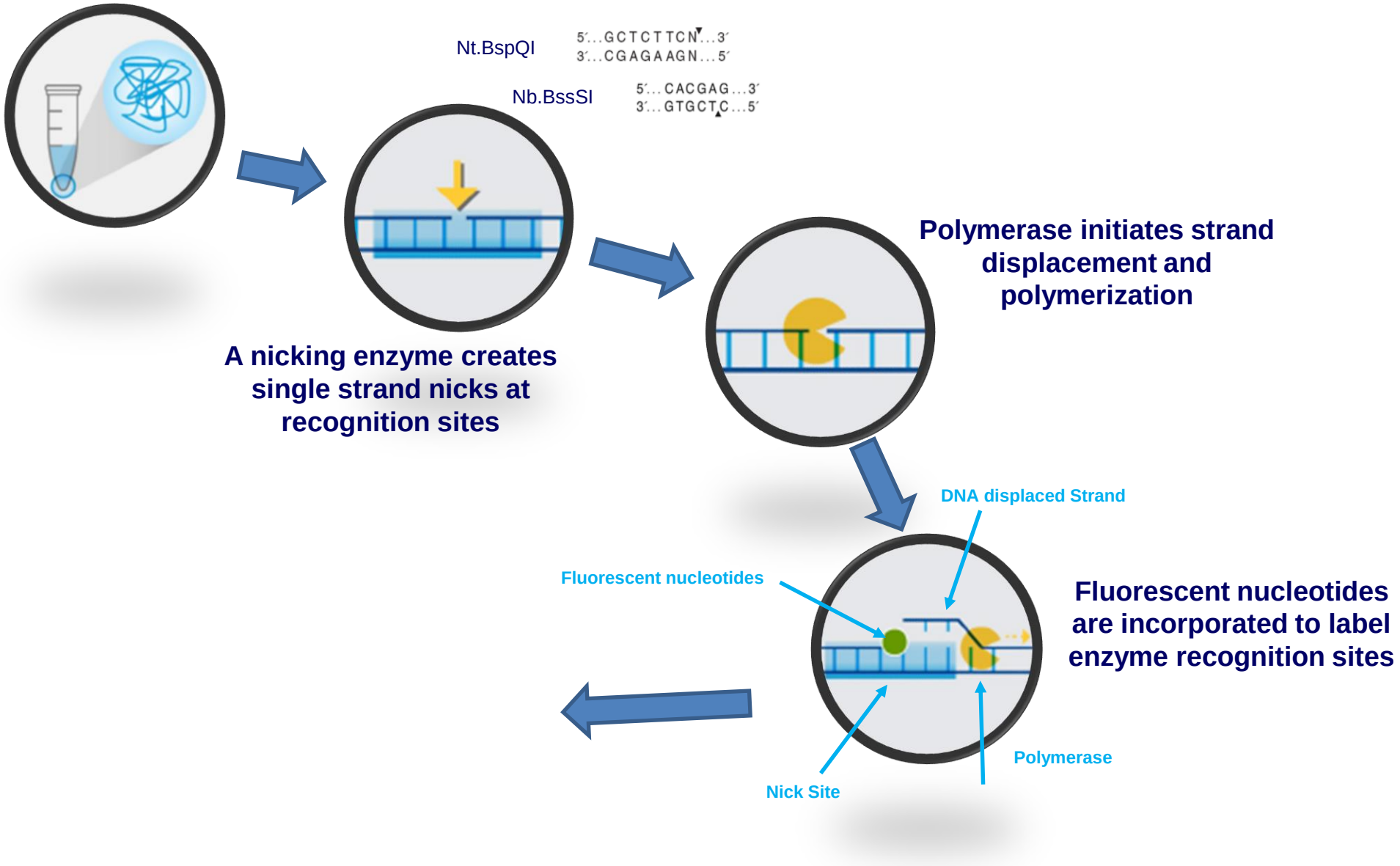


A nicking enzyme creates single strand nicks at recognition sites

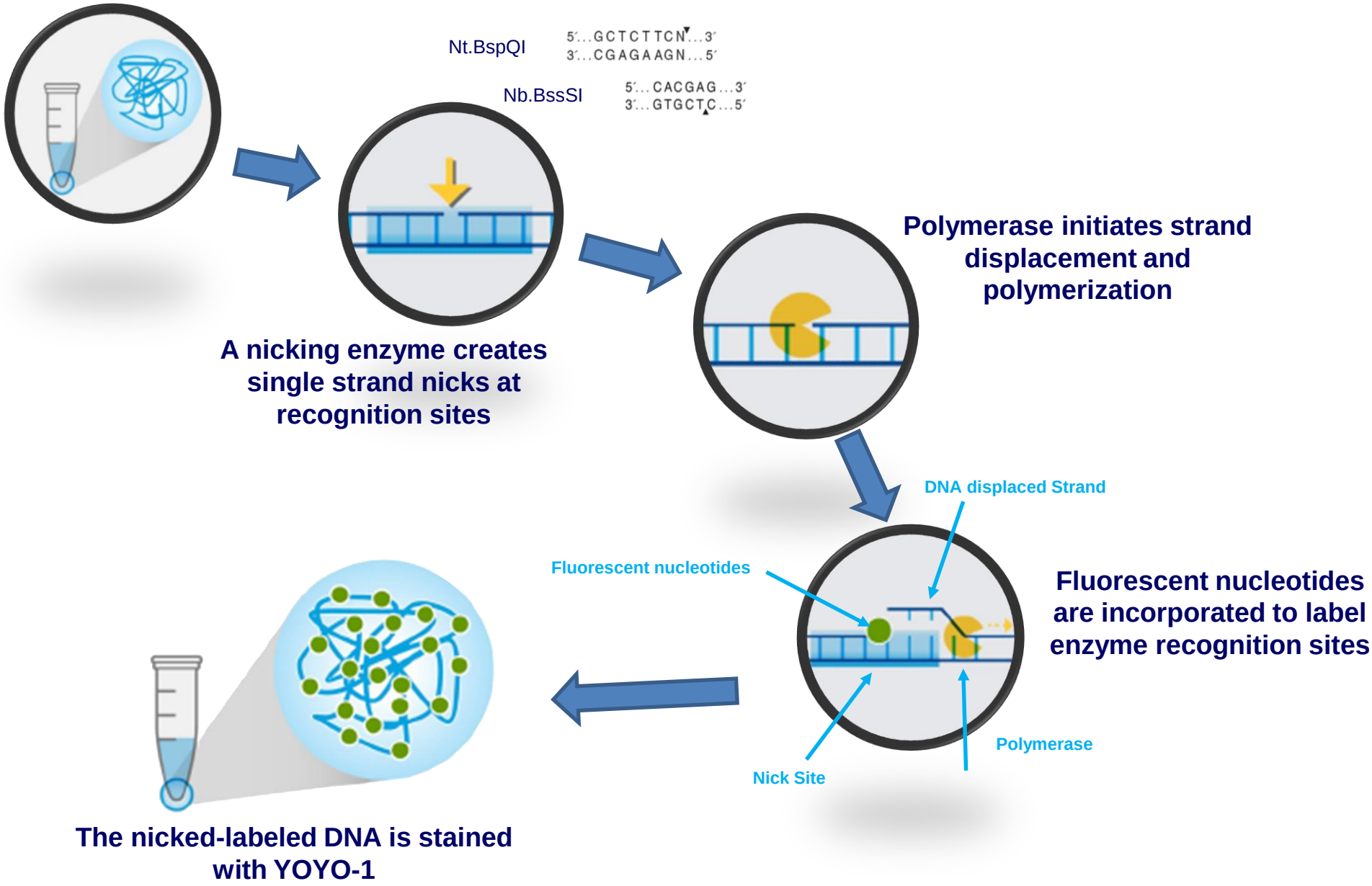


Polymerase initiates strand displacement and polymerization

Nick Label Repair Stain (NLRS) workflow



Nick Label Repair Stain (NLRS) workflow

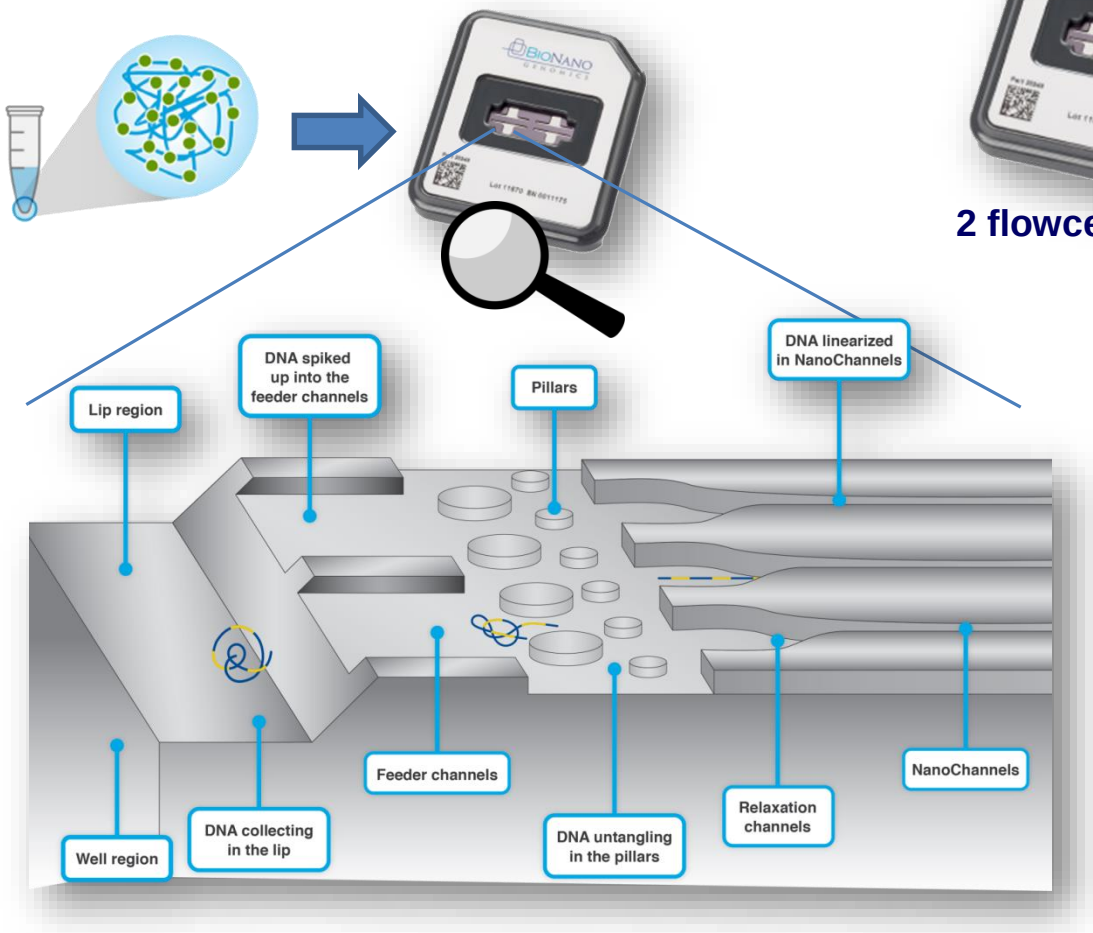


Irys System technology

The nicked-labeled-stained DNA is loaded on a chip

Irys System technology

The nicked-labeled-stained DNA is loaded on a chip

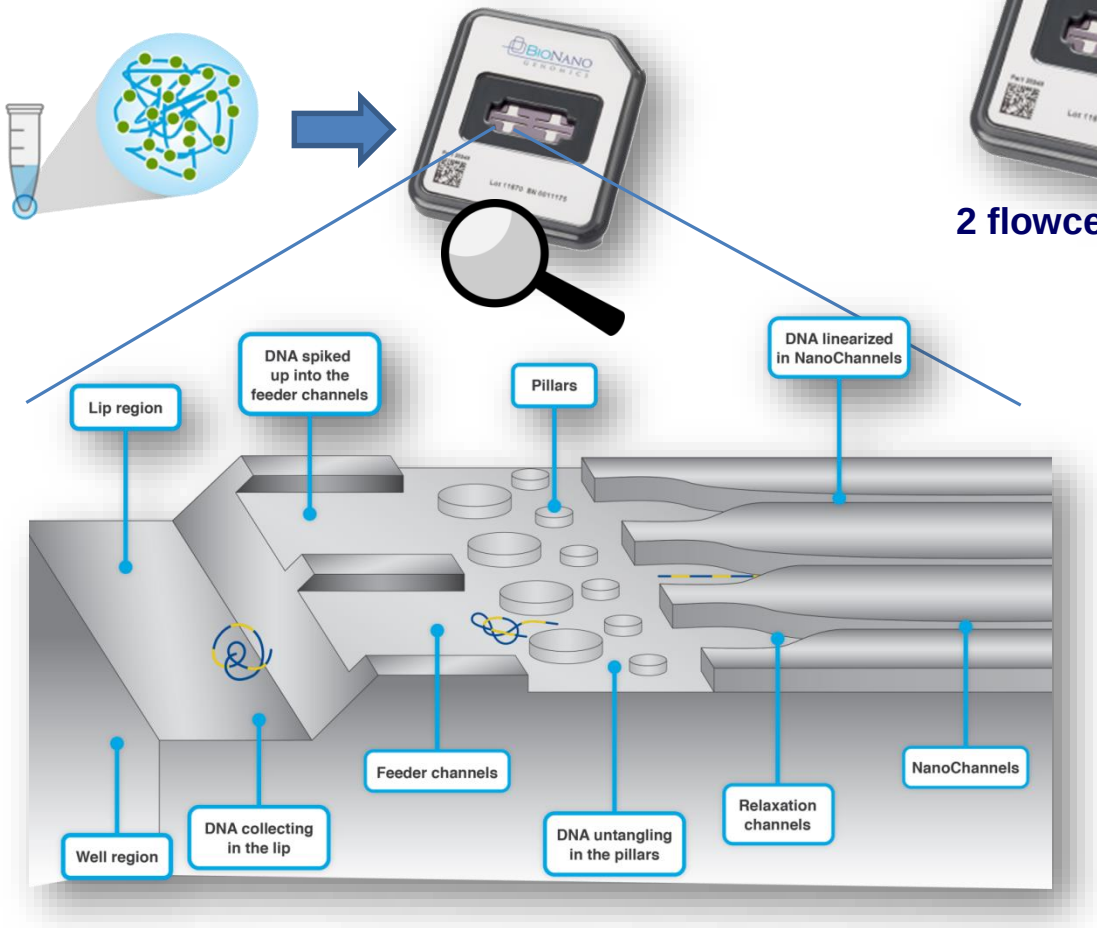


2 flowcells per Chip

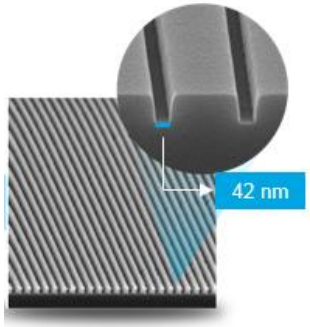
DNA molecules are linearized in the NanoChannels

Irys System technology

The nicked-labeled-stained DNA is loaded on a chip



2 flowcells per Chip



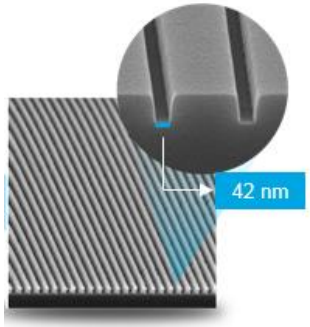
13 000 channels per flowcell

DNA molecules are linearized in the NanoChannels

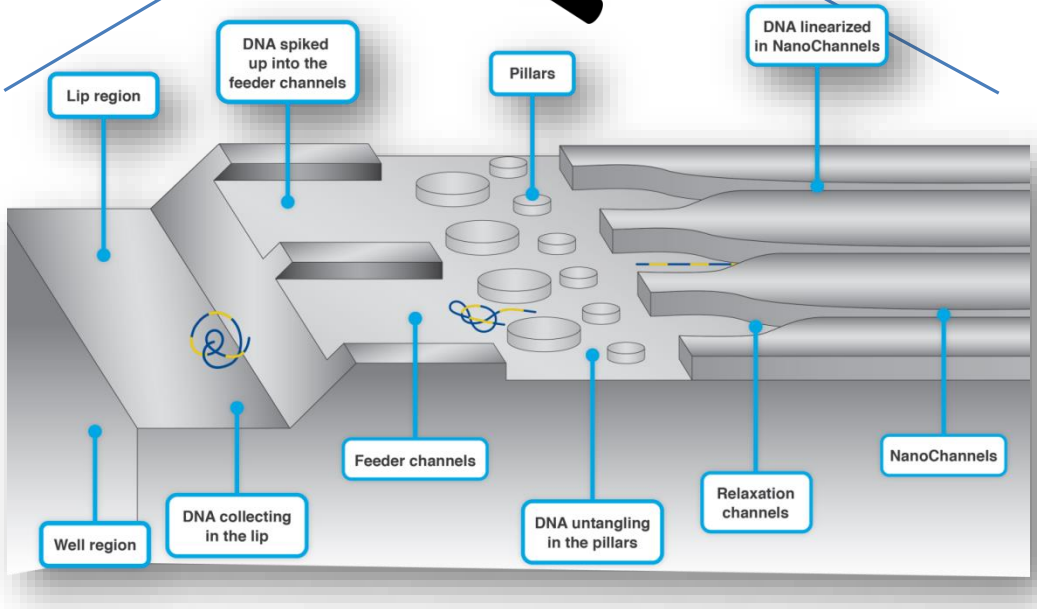
Irys System technology

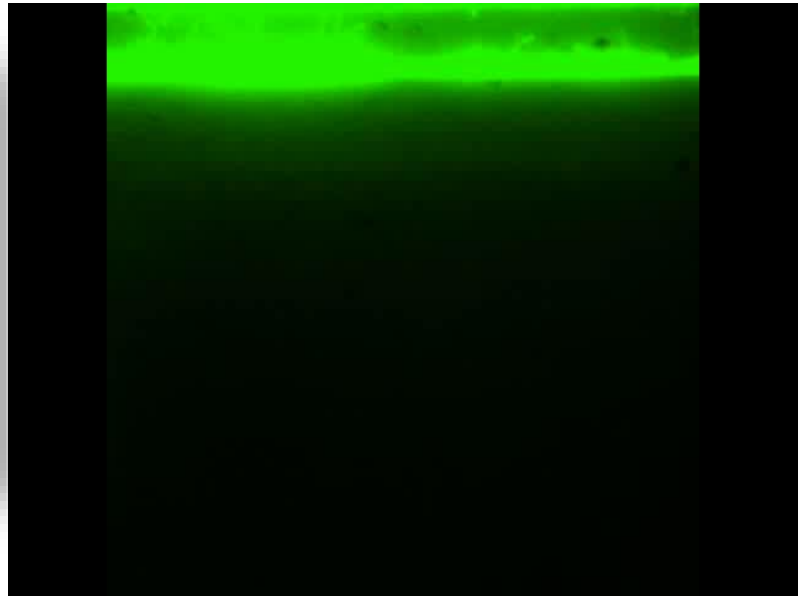
The nicked-labeled-stained DNA is loaded on a chip

2 flowcells per Chip



13 000 channels per flowcell

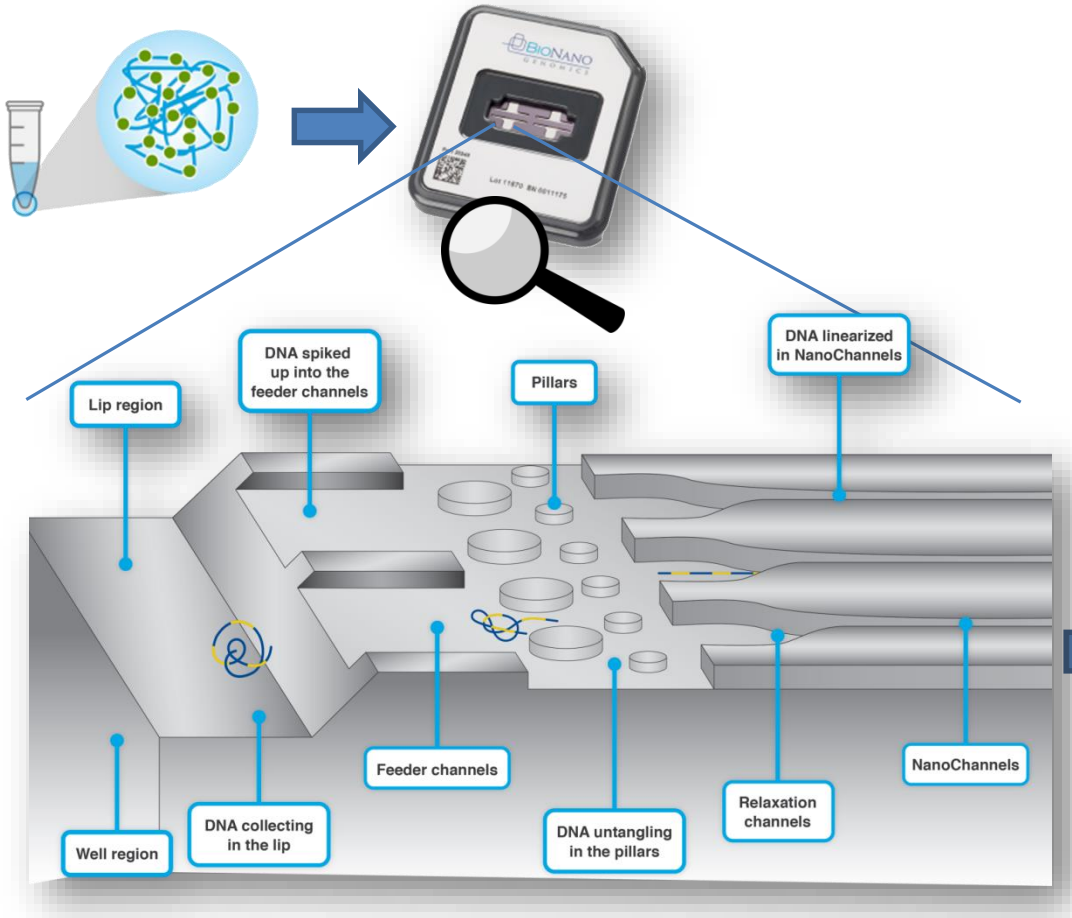




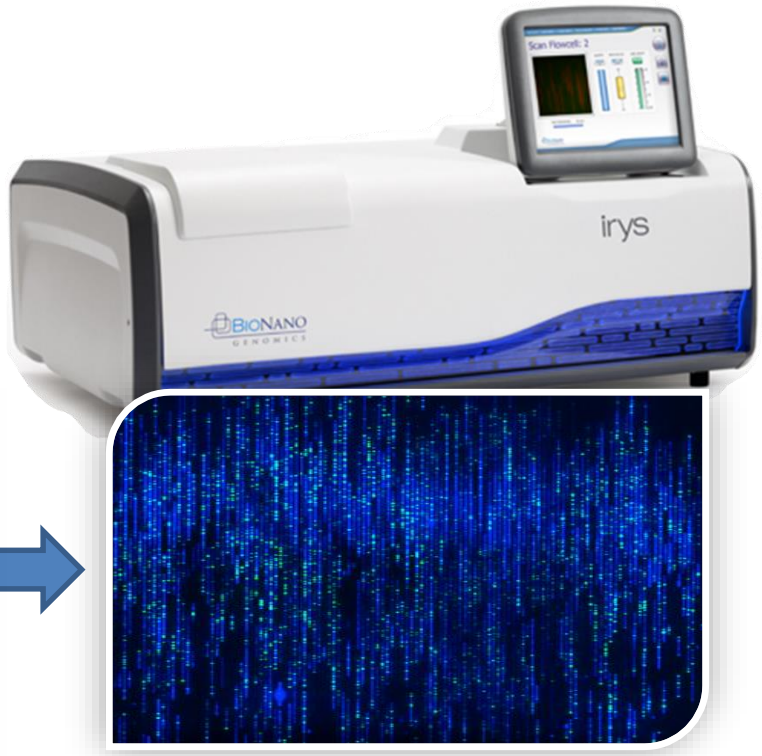
DNA molecules are linearized in the NanoChannels

Irys System technology

The nicked-labeled-stained DNA is loaded on a chip



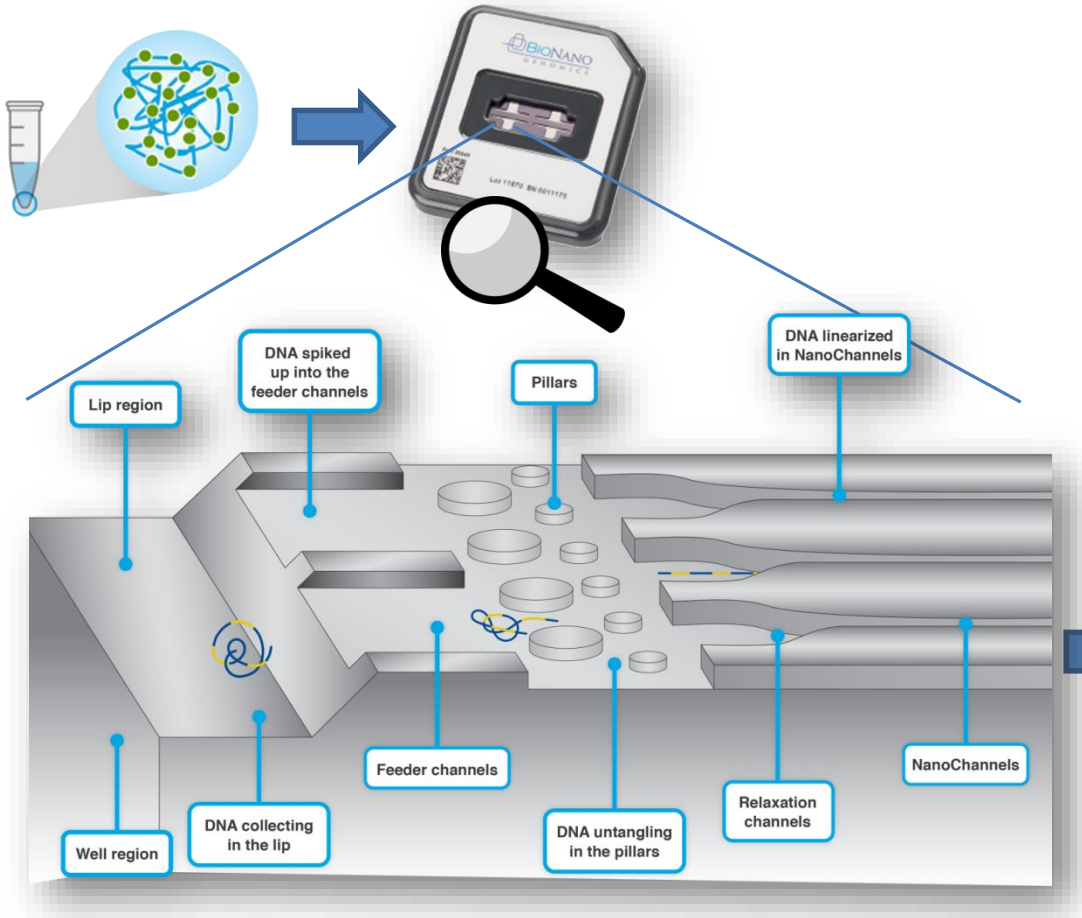
DNA molecules are linearized in the NanoChannels



The Irys System collects images of linearized DNA molecules

Irys System technology

The nicked-labeled-stained DNA is loaded on a chip

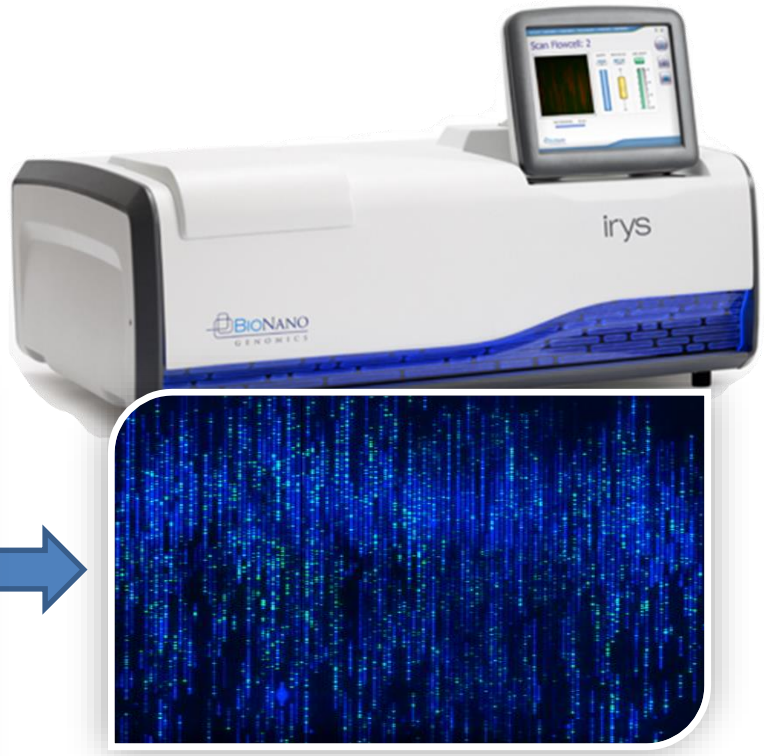


DNA molecules are linearized in the NanoChannels

30 scans/run

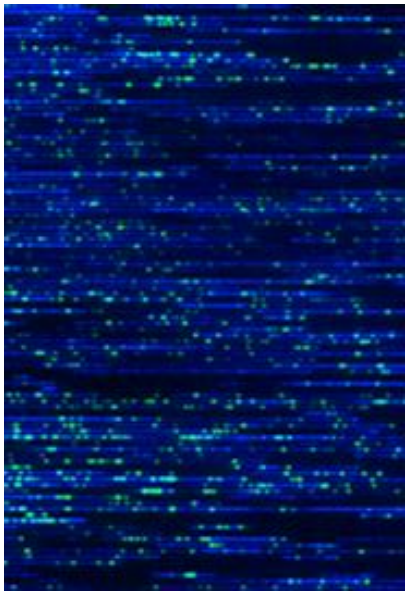
24h-30h/run

50-100 Gb raw data / run/ flowcell

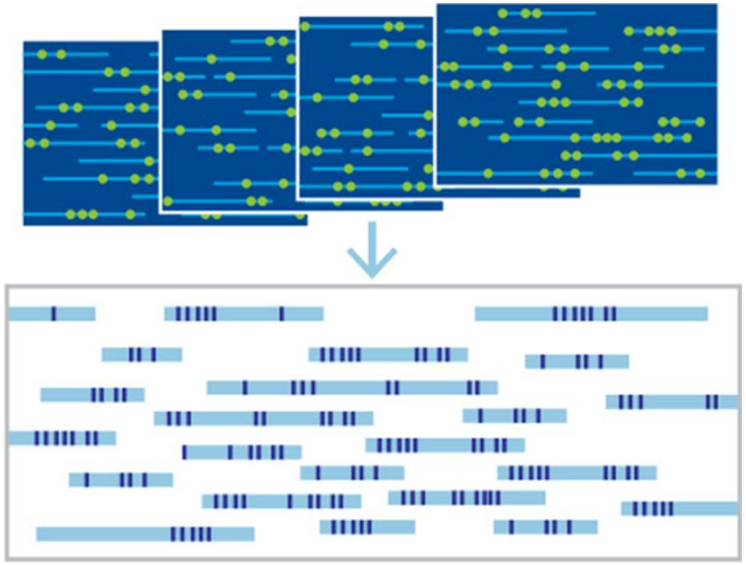


The Irys System collects images of linearized DNA molecules

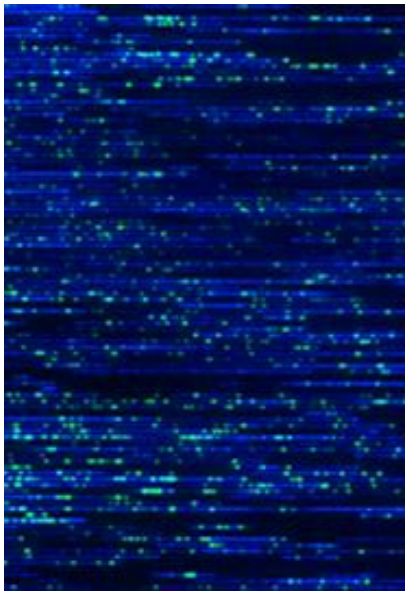
Data Analysis



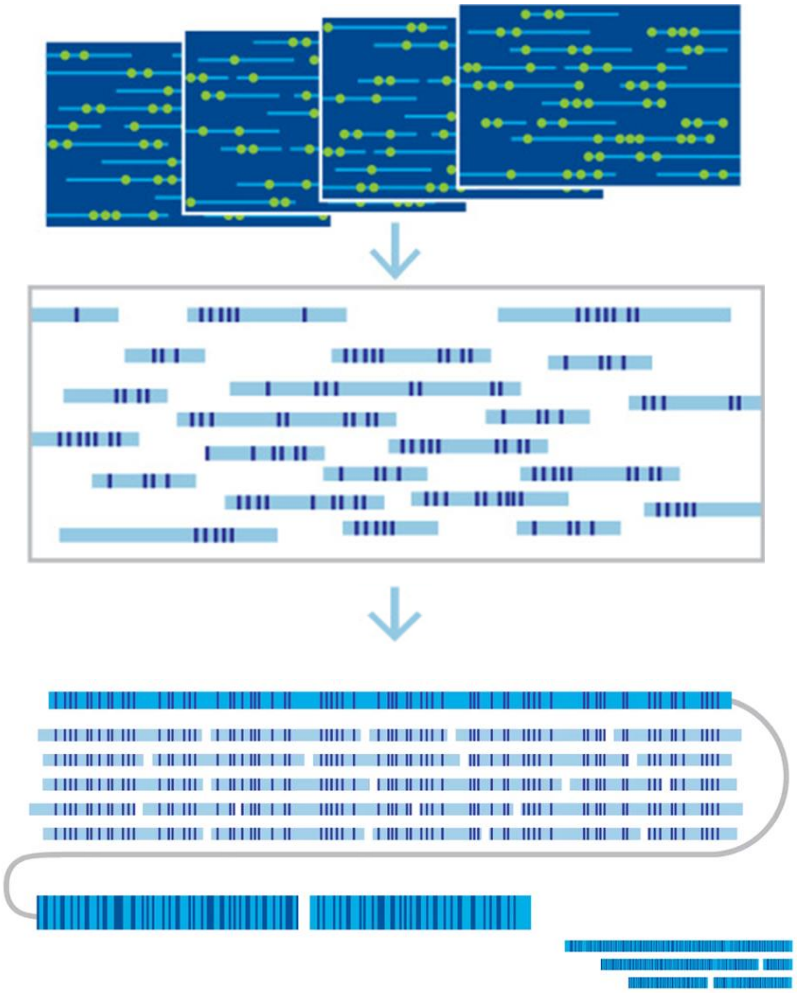
Algorithms convert images into digital molecules



Data Analysis



Algorithms convert images into digital molecules



***De Novo* assembly is made to construct a consensus genome map**

Applications

Genome Assembly and Hybrid Scaffolding

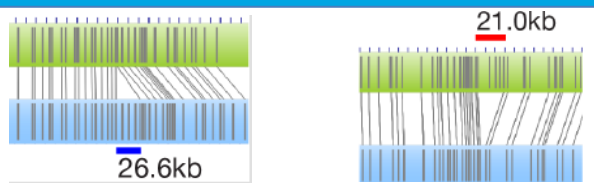
Ex: Xiao et al. 2015

- Enhance & refine genome (references and draft genomes)
- Orient and align sequencing reads using a hybrid scaffold
- Close the contig gaps
- Combined with sequencing reads for *de novo* assembly in species without a reference genome

Long range SV detection and discovery (2Kb to >1Mb)

Ex: Cao et al. 2014

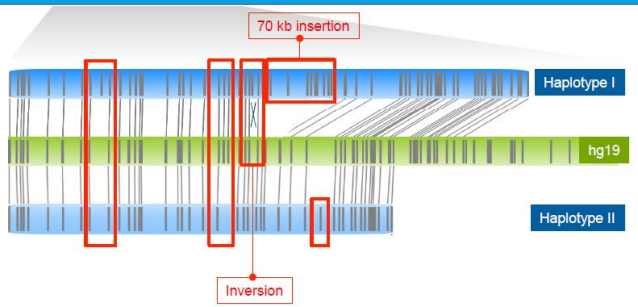
- In/Dels –Inversions
- Translocations
- CNVs



Haplotyping

Ex: Dong et al. 2016

- Comparison of 2 haplotypes with a reference

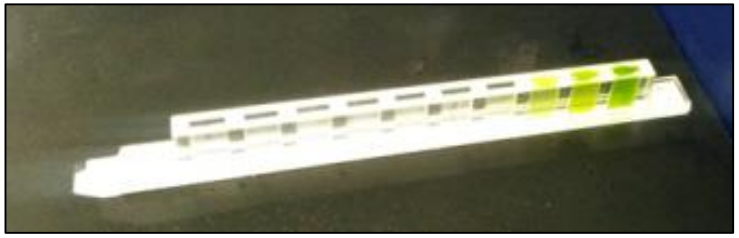


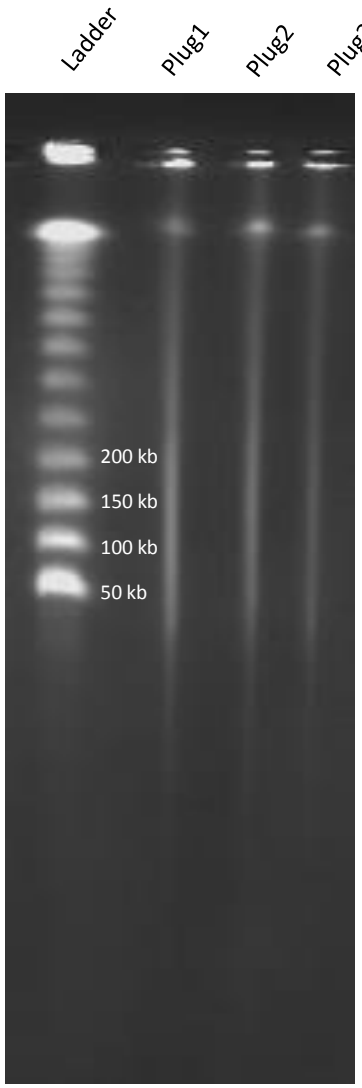
Examples of optical maps:

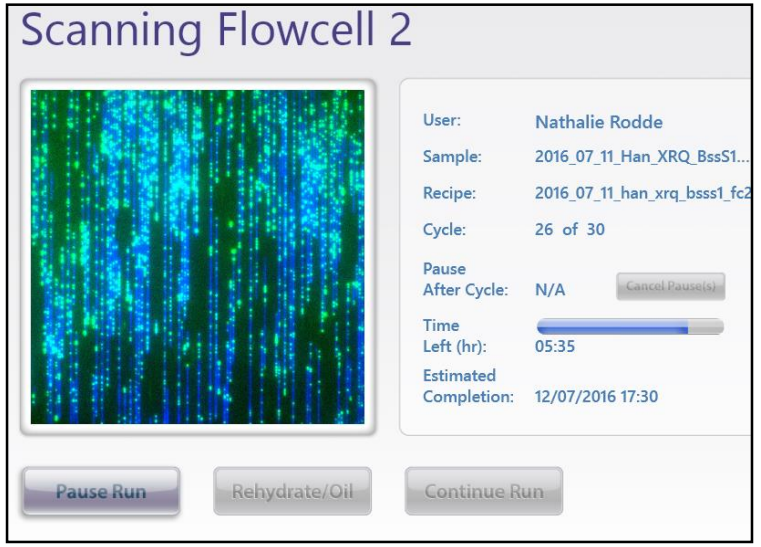
Sunflower (Helianthus *annuus*) XRQ
and
Sunflower broomrape (Orobanche *cumana*) IN-23

Example of H. *annuus* XRQ

4 g of fresh young leaves
3 days of dark treatment
2 nicking enzymes (BspQ1 & BssS1)







BspQ1
7,2 labels/100kb

1100 Gb raw data
300X
N50 168 kb

BssS1
17,2 labels/100kb

983 Gb raw data
270X
N50 148 kb

Example of H. *annuus* XRQ

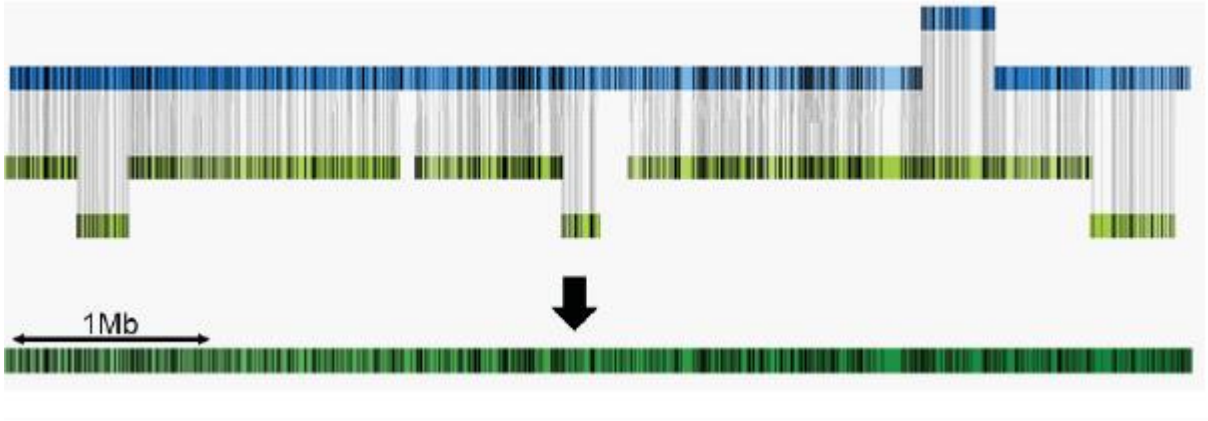
BspQ1 XRQ	PacBio assembly	BioNano *de Novo* assembly	Hybrid scaffolding (+ not scaffolded PacBio contigs)
Number of contigs	11676	2959	9433
N50 length (Mb)	0.521	1.444	2.195
Genome map length (Gb)	2.9	3.2	3.6

Estimated genome size: 3.6 Gb

BioNano Map

PacBio assembly

Hybrid scaffolding



Example of H. *annuus* XRQ

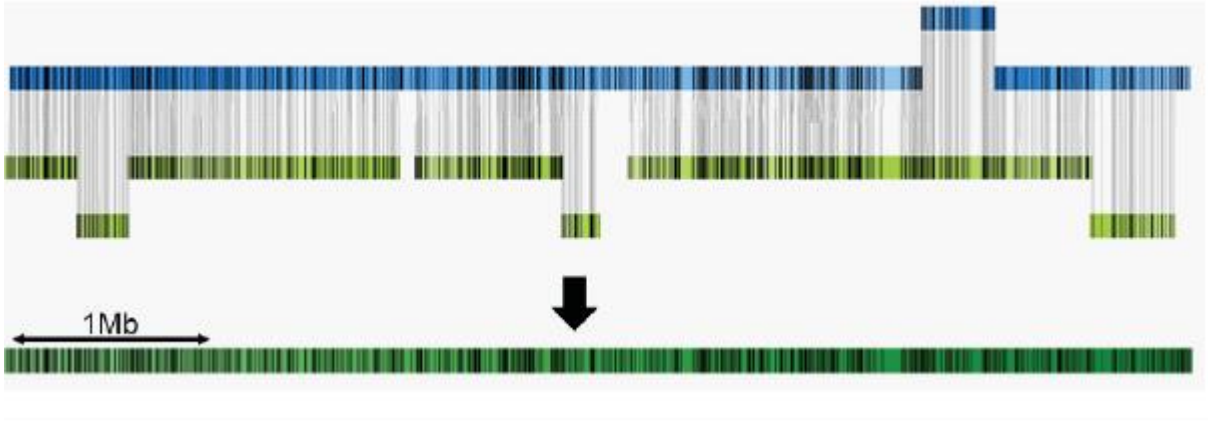
BspQ1 XRQ	PacBio assembly	BioNano *de Novo* assembly	Hybrid scaffolding (+ not scaffolded PacBio contigs)
Number of contigs	11676	2959	9433
N50 length (Mb)	0.521	1.444	2.195
Genome map length (Gb)	2.9	3.2	3.6

x 4.2

BioNano Map

PacBio assembly

Hybrid scaffolding



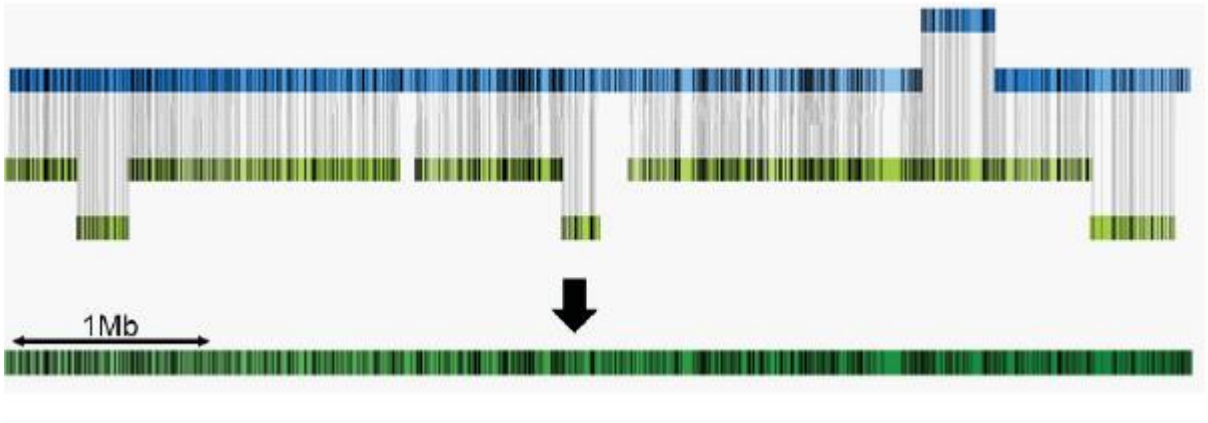
Example of H. *annuus* XRQ

Work on BssS1 data still in progress

BioNano Map

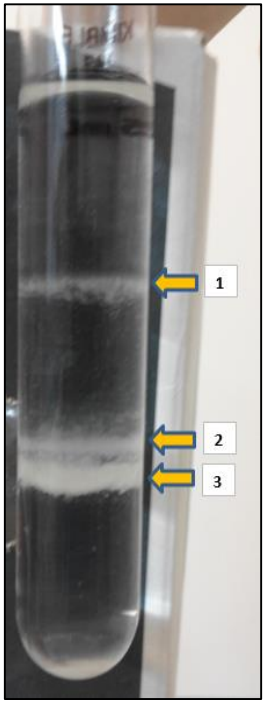
PacBio assembly

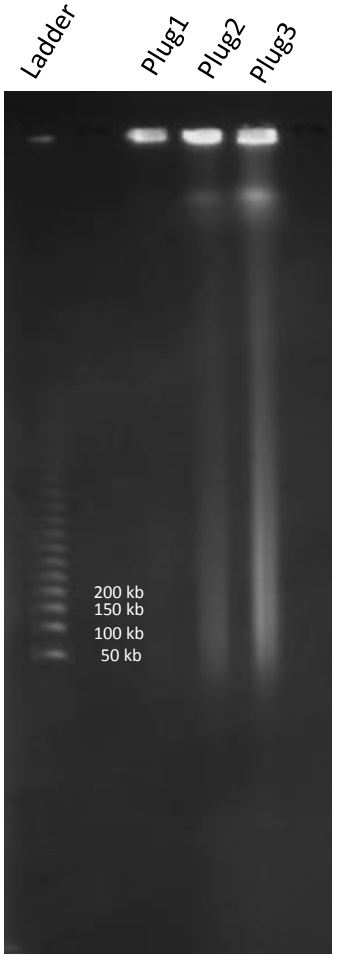
Hybrid scaffolding

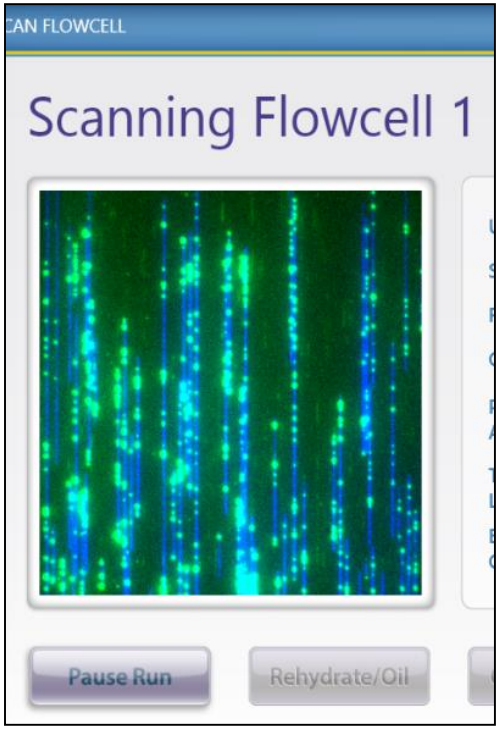


Example of O. *cumana* IN-23

2 g of frozen plant
Extraction optimization (sugar)
1 nicking enzyme (BspQ1)



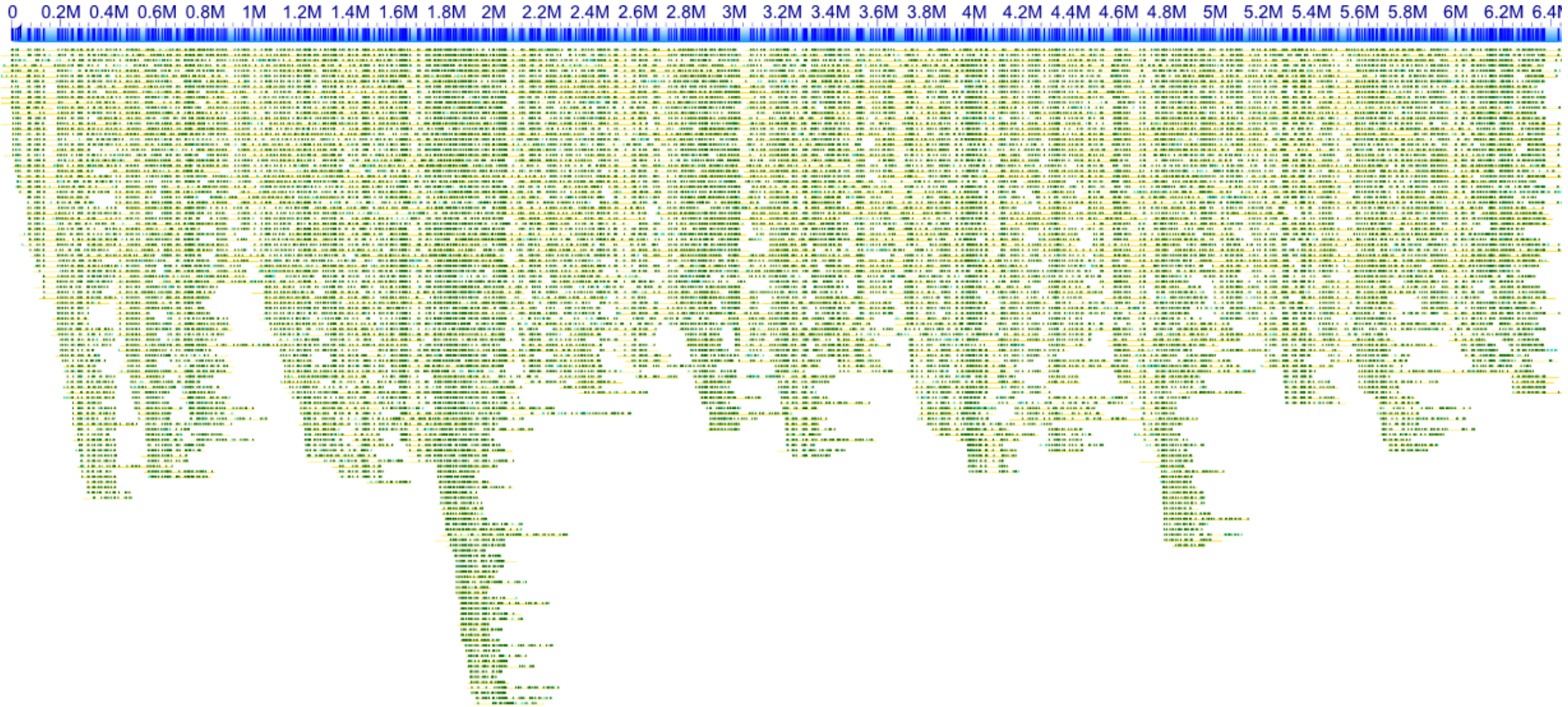




BspQ1
11,5 labels/100kb

680 Gb raw data
370X
N50 143 kb

Example of a contig assembly of *O.cumana IN-23*



Results of hybrid scaffolding of O.cumana IN-23

BspQ1 IN-23	Pacbio assembly	Bionano *de novo* assembly	Hybrid scaffolding (+ not scaffolded Pacbio contigs)
Number of Contigs	7199	1901	6405
N50 lenght (Mb)	0.901	1.204	2.389
Genome Maps Size (Gb)	1.42	1.57	1.64

Estimated genome size: 1.4 Gb – 2 Gb

Results of hybrid scaffolding of O.cumana IN-23

BspQ1 IN-23	Pacbio assembly	Bionano *de novo* assembly	Hybrid scaffolding (+ not scaffolded Pacbio contigs)
Number of Contigs	7199	1901	6405
N50 lenght (Mb)	0.901	1.204	2.389
Genome Maps Size (Gb)	1.42	1.57	1.64

Estimated genome size: 1.4 Gb – 2 Gb

Results of hybrid scaffolding of O.cumana IN-23

BspQ1 IN-23	Pacbio assembly	Bionano *de novo* assembly	Hybrid scaffolding (+ not scaffolded Pacbio contigs)
Number of Contigs	7199	1901	6405
N50 lenght (Mb)	0.901	1.204	2.389
Genome Maps Size (Gb)	1.42	1.57	1.64

x 2,6

Estimated genome size: 1.4 Gb – 2 Gb

Next steps:

- More data to improve the quality of assembly (selection of higher molecular weight DNA molecules)
- Optical map with other enzyme BssS1 (19,4 labels/100kb)
- 2 steps hybrid scaffolding

Conclusion

- **CNRGV offers genomic tools to better understand the complex genomes**

→ Construction of BAC libraries
→ Construction of optical maps

- **The optical map approach can be usefull for genome assembly/finishing (hybrid scaffolding and 2 steps hybrid scaffolding), and also allows to detect structural variations**

- **Other applications can be developped on the machine (dual labeling)**

Thank you for your attention

Céline JEZIORSKI
Arnaud BELLEC
Sonia VAUTRIN
Genséric BEYDON
Nathalie RODDE
William MARANDE
Joëlle FOURMENT
Elisa PRAT
Nadine GAUTIER
Nadège ARNAL
Céline CHANTRY-DARMON
Stéphane CAUET
David PUJOL
Laetitia HOARAU
Hélène BERGES

Laboratoire Interactions Plantes Micro-organismes

Jérôme GOUZY
Marie-Claude BONIFACE
Johann LOUARN
Nicolas POUILLY
Nicolas LANGLADE
Stéphane MUNOS

http://cnrgv.toulouse.inra.fr/

@CNRGV