Toward a better understanding of plant genomes structure: combining NGS and optical mapping technology to improve the sunflower assembly

Céline CHANTRY-DARMON

INRA SCIENCE & IMPACT
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

CNRGV
The French Plant Genomic Center

- Created in 2004 by INRA
- A dedicated structure to assist plant genomic programs
  - ➢ Distribute the genomic resources at the international level
  - ➢ Provide high quality research material and efficient tools and services
  - ➢ Develop genomic projects in collaboration
  - ➢ Host scientists
  - ➢ Develop innovative solutions

ISO 9001:2008
Octobre 2005

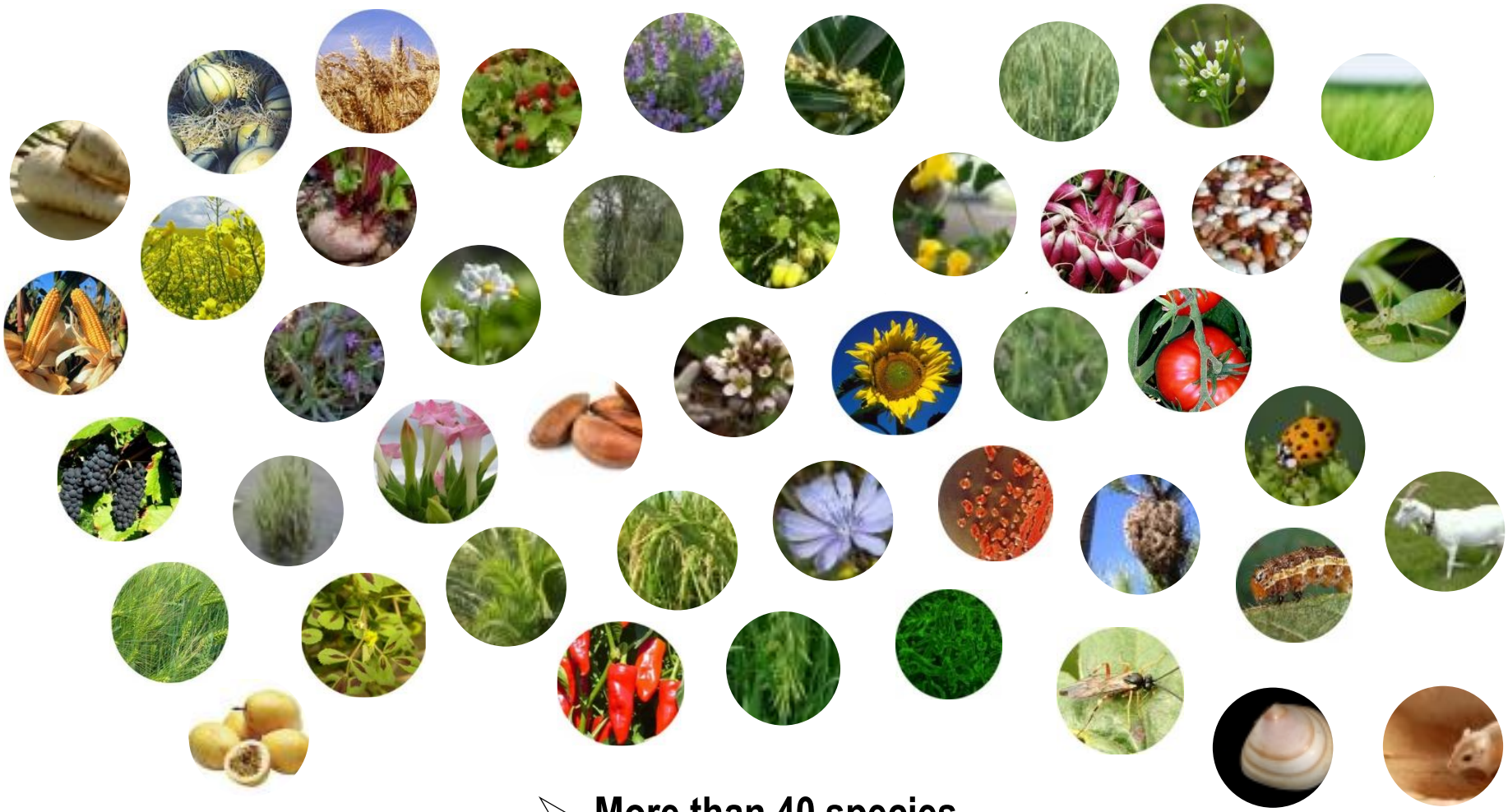
INRA SCIENCE & IMPACT
CIRAD
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Interactions with laboratories around the world

- More than 3 millions BAC clones distributed during the last 5 years

INRA SCIENCE & IMPACT
CNRGV
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

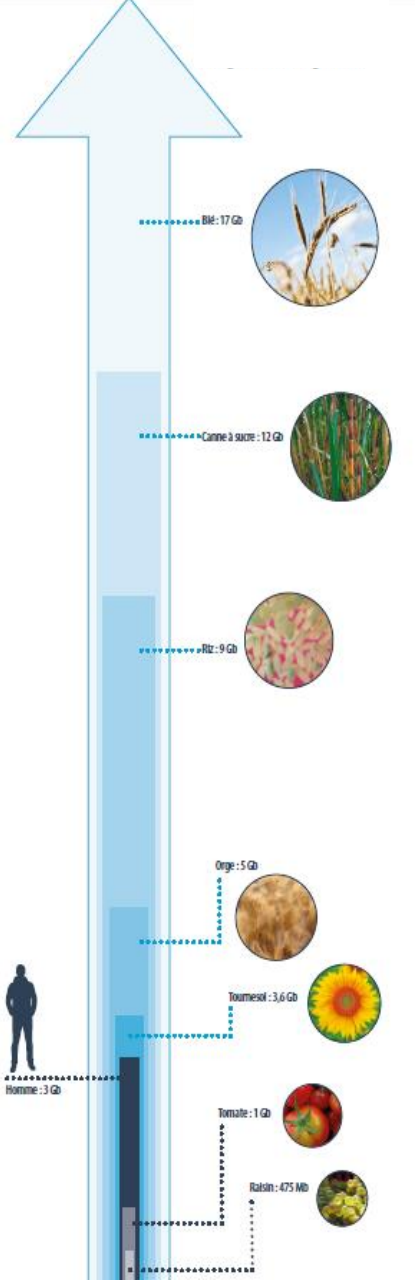
Plants project diversity



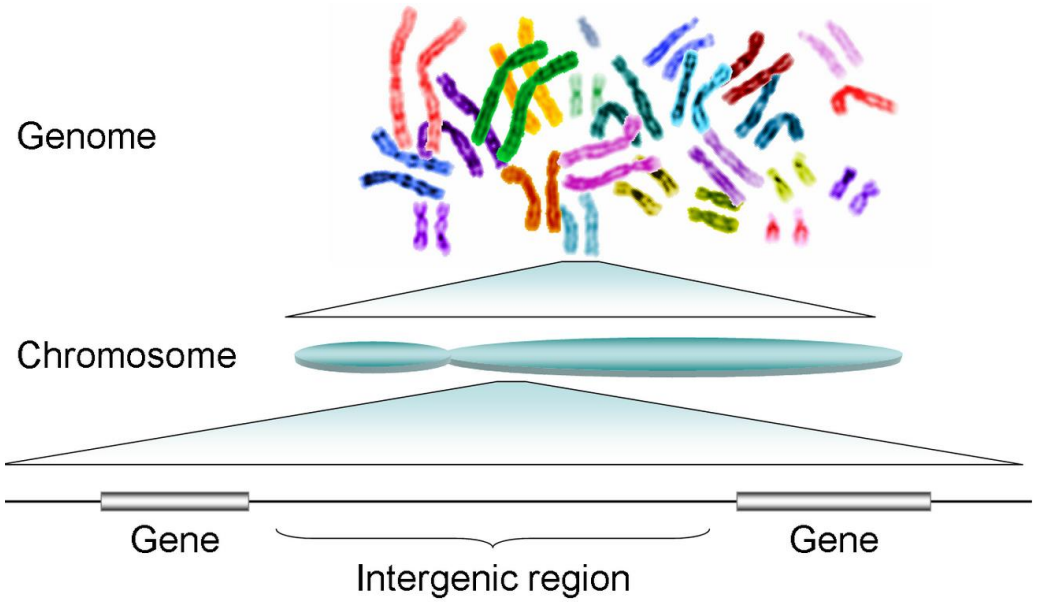
- More than 40 species

INRA SCIENCE & IMPACT
CNRGV
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

The goal for the Pant Genomic Center



- Large genome size
- Repeats elements
- Polyploidy

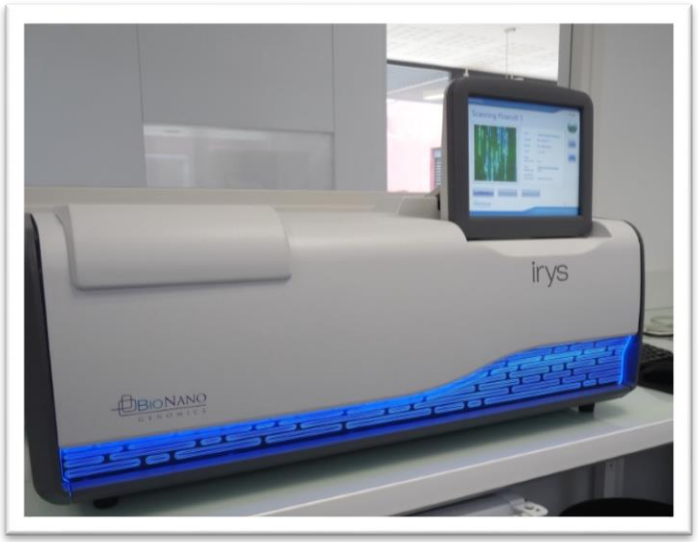


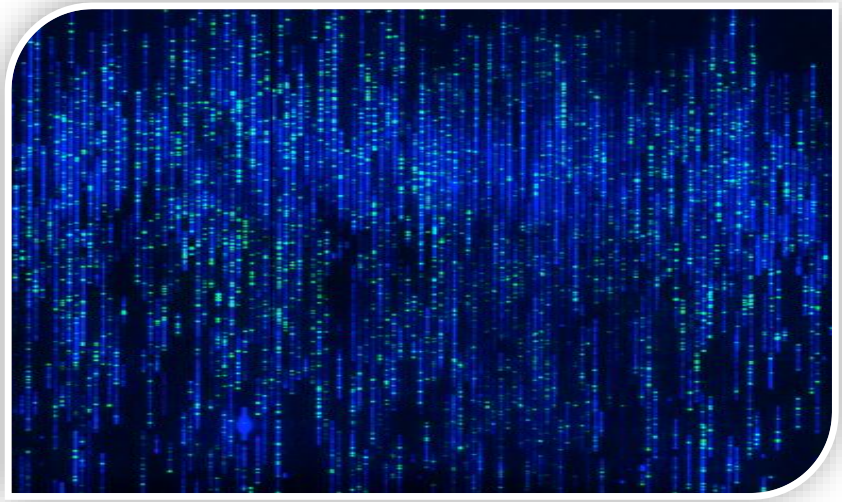
- ➢ Manage genome size and diversity
- ➢ Decrease genome complexity
- ➢ Target genomic region of interest

INRA SCIENCE & IMPACT
CNRGV
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Focus on the optical mapping with BioNano

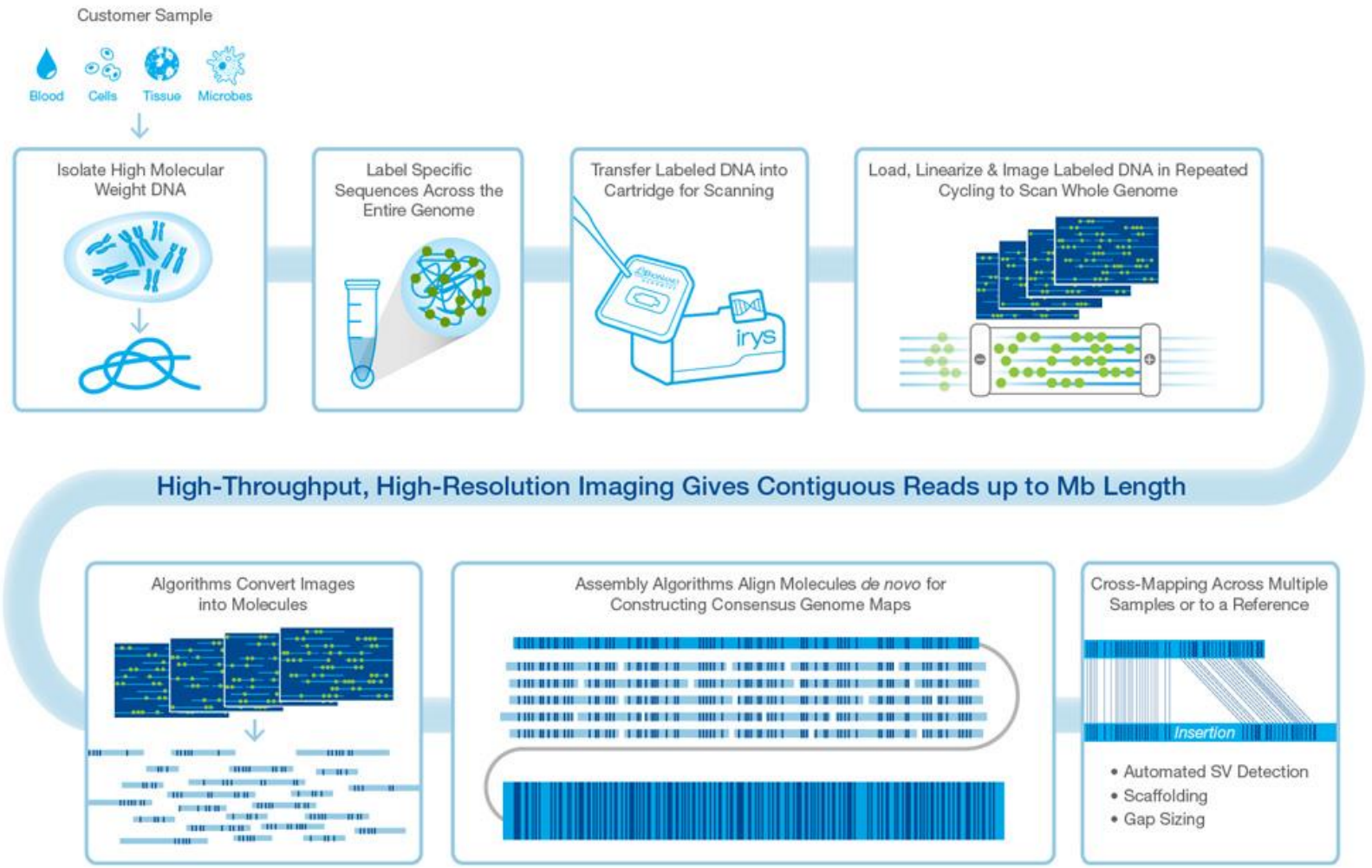
- The Bionano Irys system: a new tool to study complex genomes
- Advantages of BioNano optical mapping:
  - Direct visualization of long DNA molecules (>100 kb)
  - Provides real physical distance information





CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

The Workflow

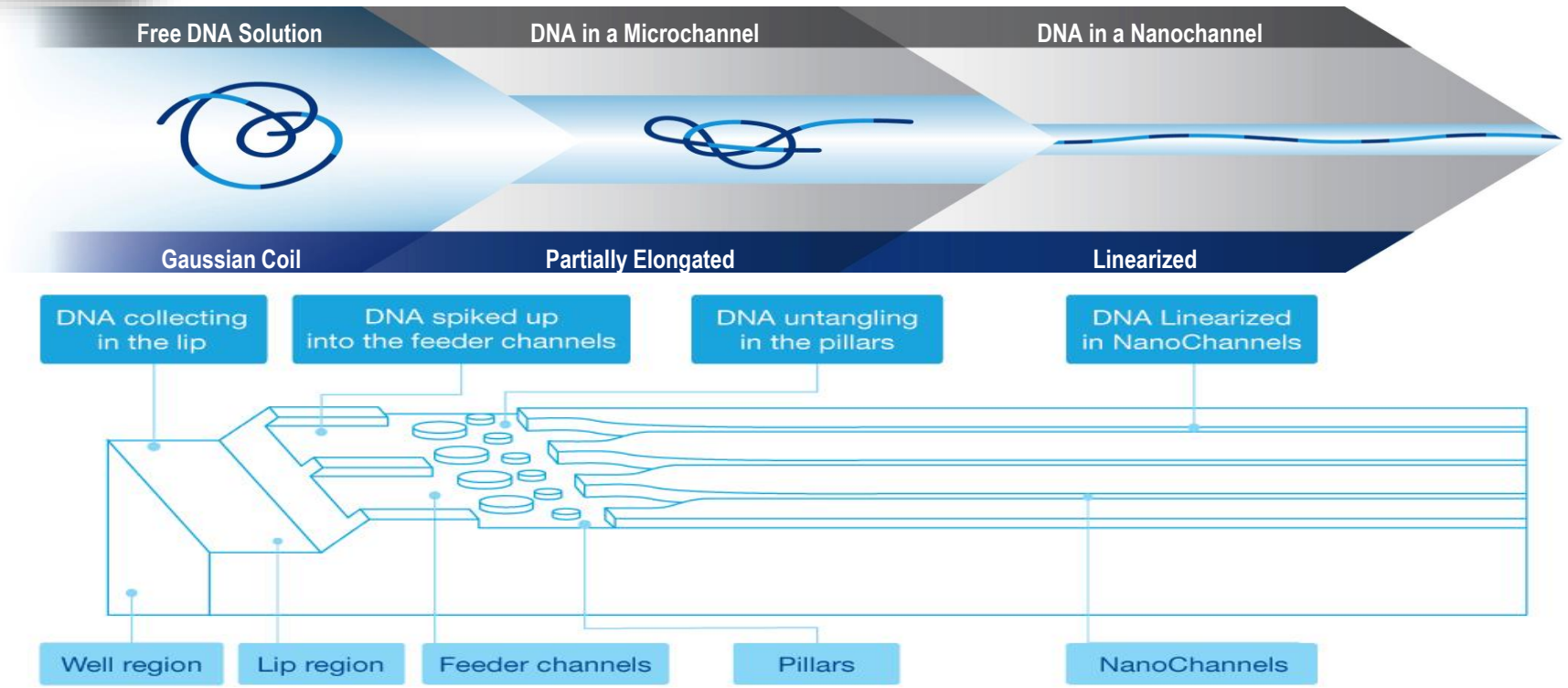


➢ **50Gb data generated per flowcell (=> 100Gb / chip)**

BIONANO GENOMICS

CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

HMW DNA molecules in nanochanels

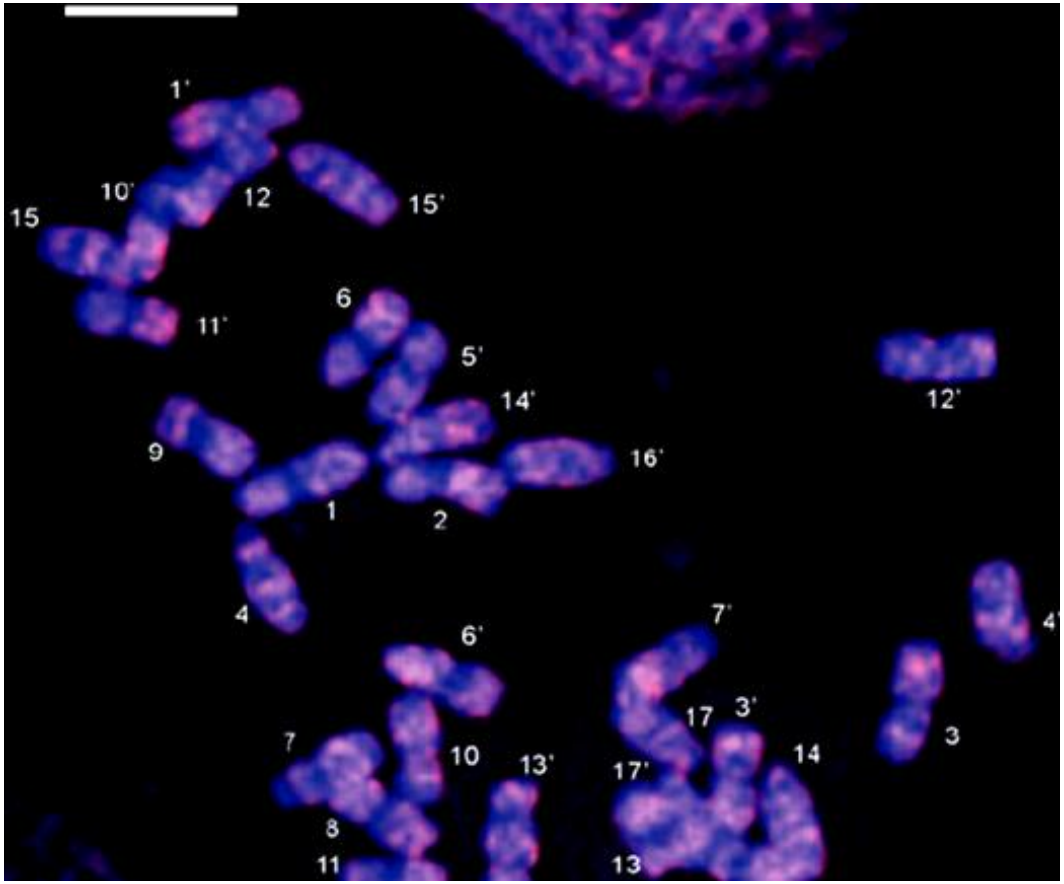


- **HMW DNA molecules from 100kb to >2Mb**

BIONANO GENOMICS

CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

The Sunflower genome



- *Helianthus annuus*
- 3.6 Gb
- 2n=34 chromosomes

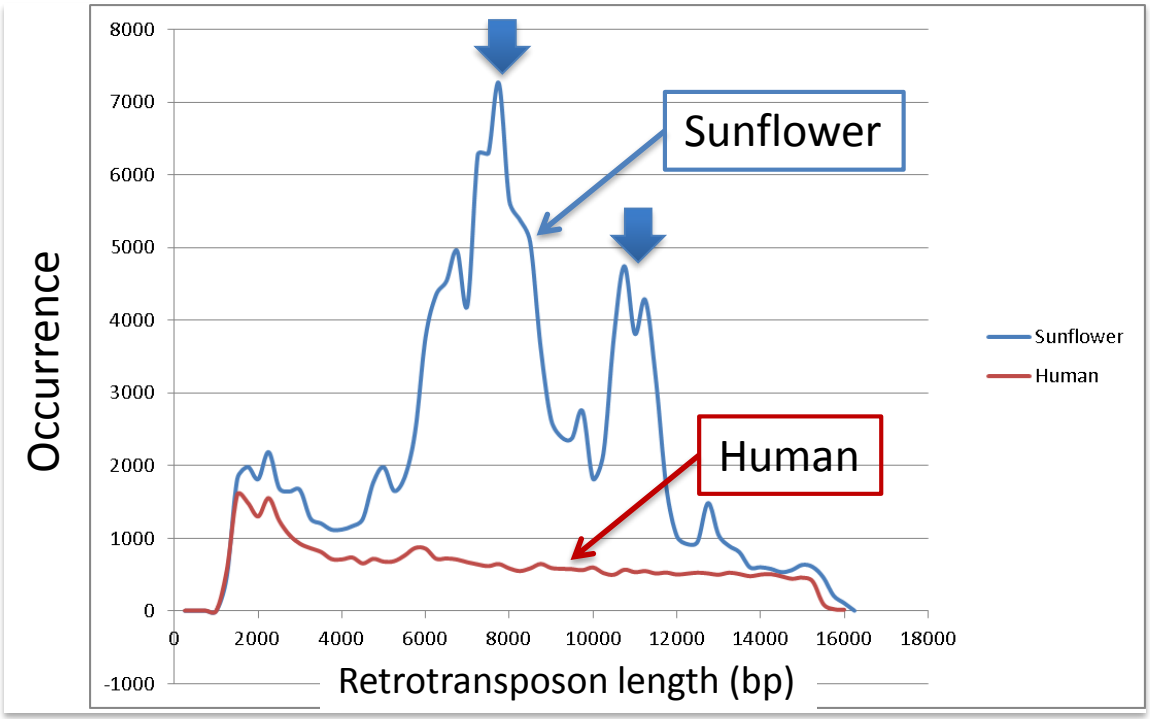
Cytological characterization of sunflower by in situ hybridization using homologous rDNA sequences and a BAC clone containing highly represented repetitive retrotransposon-like sequences

P. Talia, E. Greizerstein, C. Díaz Quijano, L. Peluffo, L. Fernández, P. Fernández, H.E. Hopp, N. Paniego, R.A. Heinz, and L. Poggio

CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Sunflower genome contains long repeated sequences

Length distribution of LTR retrotransposons



LTRharvest (Ellinghaus *et al.* 2008, default parameters)

J. Gouzy

Repeats = 33% of the sunflower genome

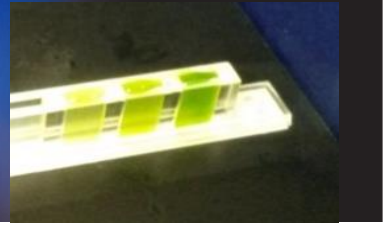
Repeats = 8% of the Human genome

Two major repeats in the sunflower genome:
8 kb and 11.5 kb

The repeats make the assembling very difficult

INRA SCIENCE & IMPACT
CNRGV PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

BioNano analyses

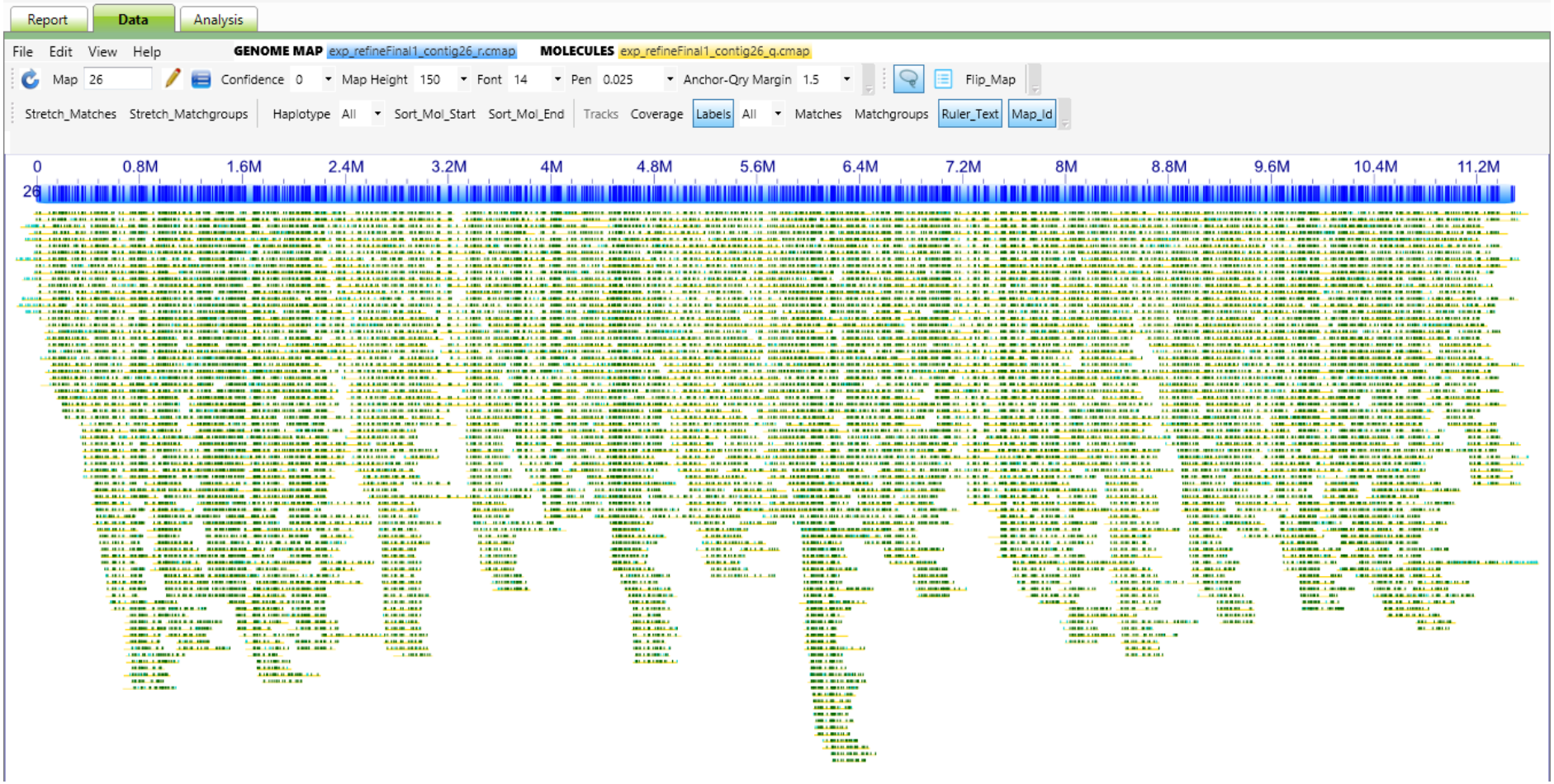


- HMW DNA Extraction of fresh young dark treated leaves
- 2 nicking enzymes (BspQ1 & BssS1)

	BspQ1	Bsss1
	5′...GCTCTTCN▼...3′ 3′...CGAGAAGN...5′	5′...CACGAG...3′ 3′...GTGCTC▲...5′
Theoretical nb labels / 100kb	7,2	17,2
Real nb labels / 100kb	6,4	12,8
Raw data (Gb)	846 (235X)	845 (235X)
Filtered data >100kb (Gb)	635 (176X)	600 (167X)
Molecules N50 (kb)	206	187

CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

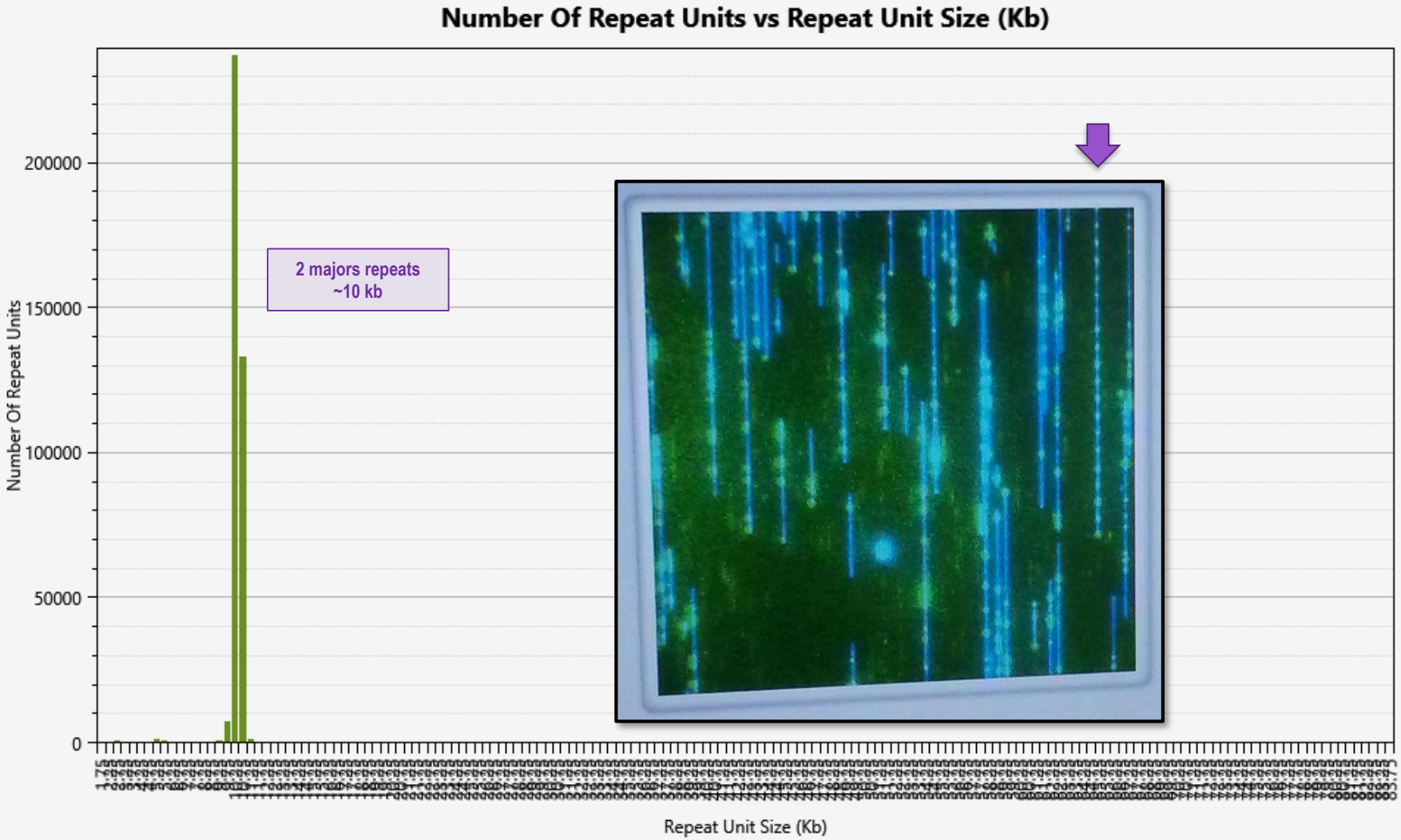
Example of a BioNano map



- 176X coverage, molecules from 150kb to 2,3Mb

INRA SCIENCE & IMPACT
CNRGV
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Visualization of the repeats



INRA SCIENCE & IMPACT
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

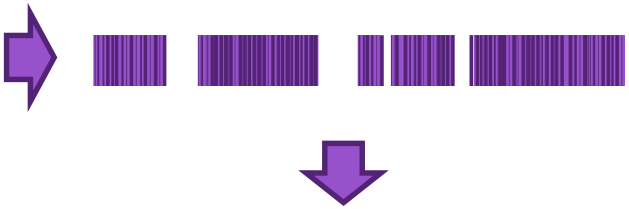
BioNano Maps

	PacBio Assembly	BioNano BspQ1 Assembly	BioNano BssS1 Assembly
Count	12318	2228	4287
Median length (Mb)	0.120	0.999	0.551
N50 length (Mb)	0.524	1.979	0.968
Max length (Mb)	3.35	11.49	5.322
Total length (Mb)	2930	3191	3112
% genome coverage	81%	88%	86%

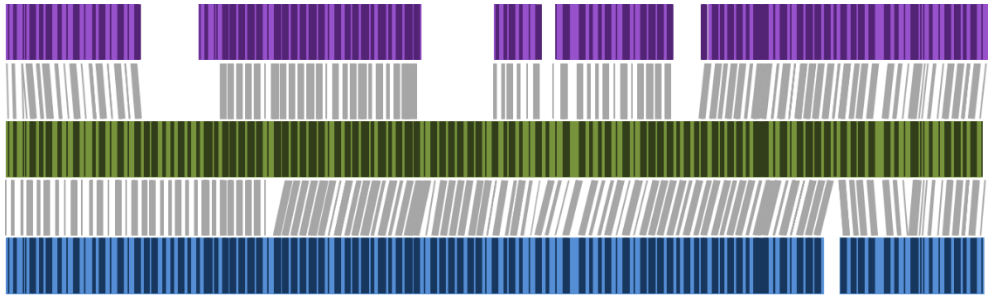
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Hybrid Assembly

PacBio Assembly

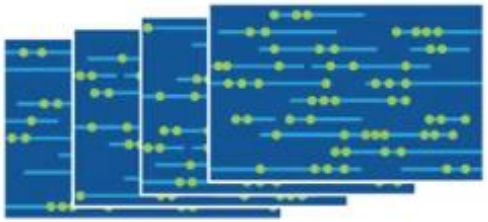


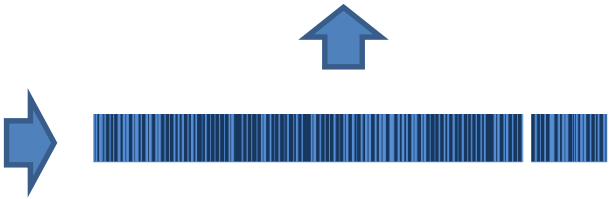
Nickase sites on NGS contigs



Hybrid scaffold

BioNano Mapping





Nickase event Fingerprint

INRA SCIENCE & IMPACT
PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Sunflower Hybrid Assembly

	PacBio Assembly	BioNano BspQ1 Assembly	Hybrid scaffold
Count	12318	2228	1430
Median length (Mb)	0.120	0.999	1.442
N50 length (Mb)	0.524	1.979	2.87
Max length (Mb)	3.35	11.49	17.45
Total length (Mb)	2930	3191	2922
% genome	81%	88%	81%

More than 5 fold increase

INRA SCIENCE & IMPACT
CNRGV PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Sunflower genome enhancement

	PacBio Assembly	BioNano BspQ1 Assembly	Hybrid scaffold
Count	12318	2228	1430
Median length (Mb)	0.120	0.999	1.442
N50 length (Mb)	0.524	1.979	2.87
Max length (Mb)	3.35	11.49	17.45
Total length (Mb)	2930	3191	2922
% genome	81%	88%	81%

Hybrid scaffold + not scaffolded PacBio contigs : 3611Mb

CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

2 Step Hybrid Assembly

	PacBio Assembly	BioNano BspQ1 Assembly	Hybrid scaffold BspQ1	BioNano BssS1 Assembly	Hybrid scaffold 2 Step
Count	12318	2228	1430	4287	1069
Median length (Mb)	0.120	0.999	1.442	0.551	1.914
N50 length (Mb)	0.524	1.979	2.87	0.968	4.166
Max length (Mb)	3.35	11.49	17.45	5.322	24.670
Total length (Mb)	2930	3191	2922	3112	2960
% genome	81%	88%	81%	86%	82%

More than 7 fold increase

INRA SCIENCE & IMPACT
CNRGV PLANT GENOMIC CENTER
CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Conclusion

- This preliminary results are enhancing the sunflower genome
- More than 7 fold improvement of the N50 length
- 2-step hybrid scaffolding strategy improves significantly the resulting N50
- Now we must look to the conflicts between the NGS assembly and the BN assembly more in details

CNRGV-INRA
24 chemin de Borde Rouge/ C.S. 52627 31326 Castanet-Tolosan
Tél: 05 61 28 52 53 / Fax: 05 61 28 55 64

Acknowledgements

@CNRGV
@SUNRISE_France

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

Jérôme GOUZY
Nicolas LANGLADE
Stéphane MUNOS

John BAETEN
Kees-Jan FRANCOIJS

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