SUNRISE
UNE CULTURE POUR LE FUTUR

SUNRISE
UNE CULTURE POUR LE FUTUR

Utilisation des données de séquençage de ressources phytogénétiques
Panorama actuel et enjeux
Exemple du tournesol

Jérôme Gouzy, Nicolas Langlade

INRA Toulouse

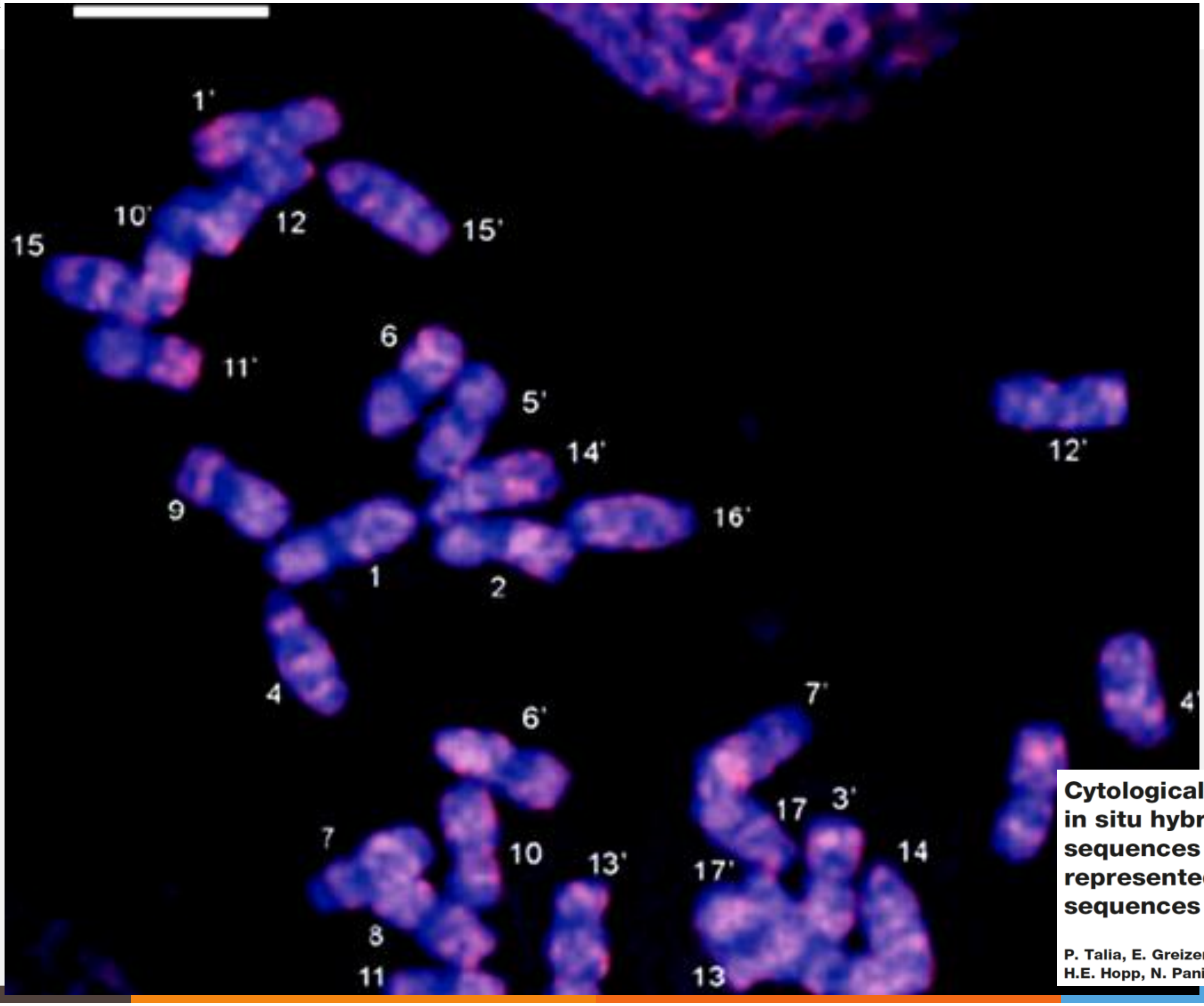
Loren H. Rieseberg

UBC Vancouver

John M. Burke

UGA Athens

Sunflower genome background

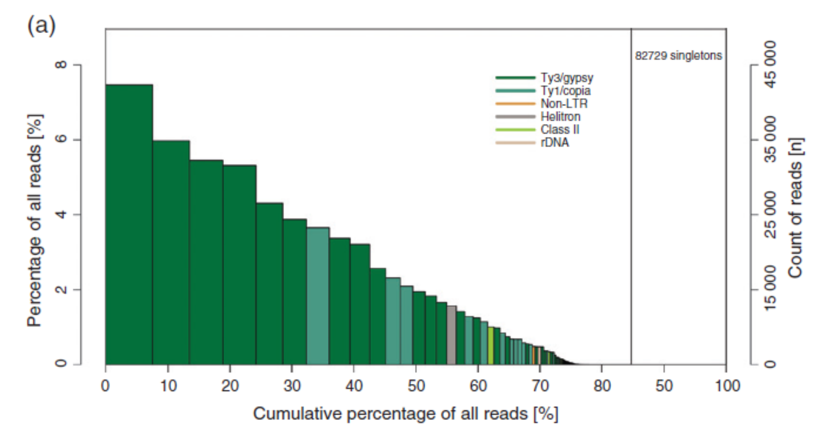


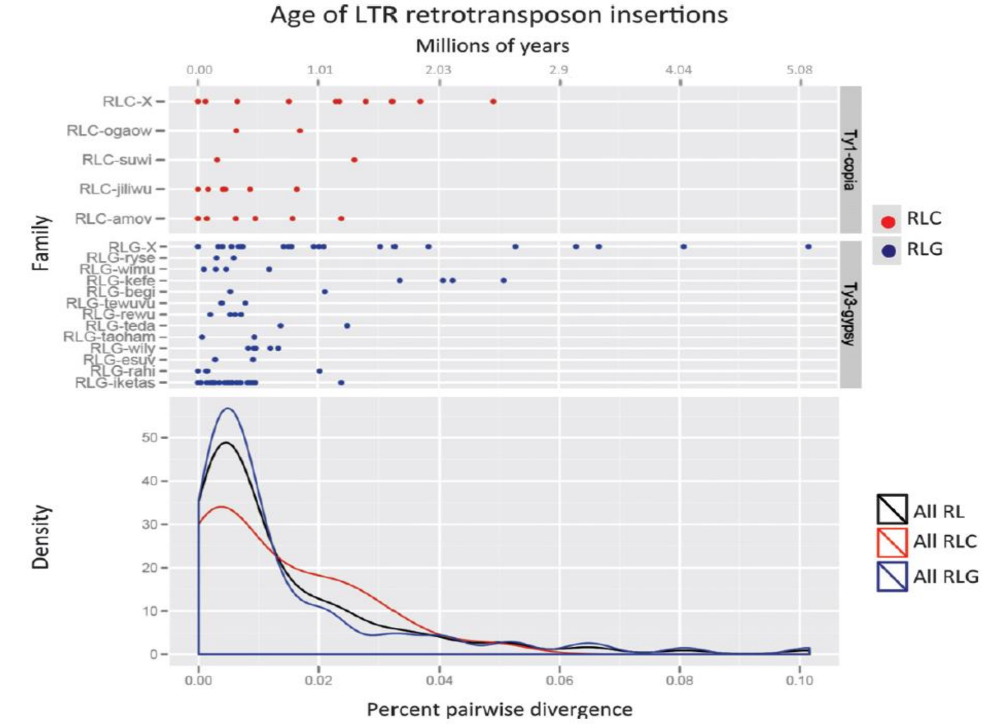
Species	Size
Rice	430 Mb
Rapeseed	1 100 Mb
Maize	2 300 Mb
H. sapiens	3 200 Mb
Sunflower	**3 600 Mb**
Wheat	17 000 Mb

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

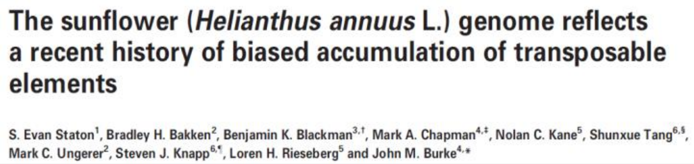
P. Talia, E. Greizerstein, C. Díaz Quijano, L. Peluffo, L. Fernández, P. Fernández, H.E. Hopp, N. Paniego, R.A. Heinz, and L. Poggio

Sunflower genome is highly repeated





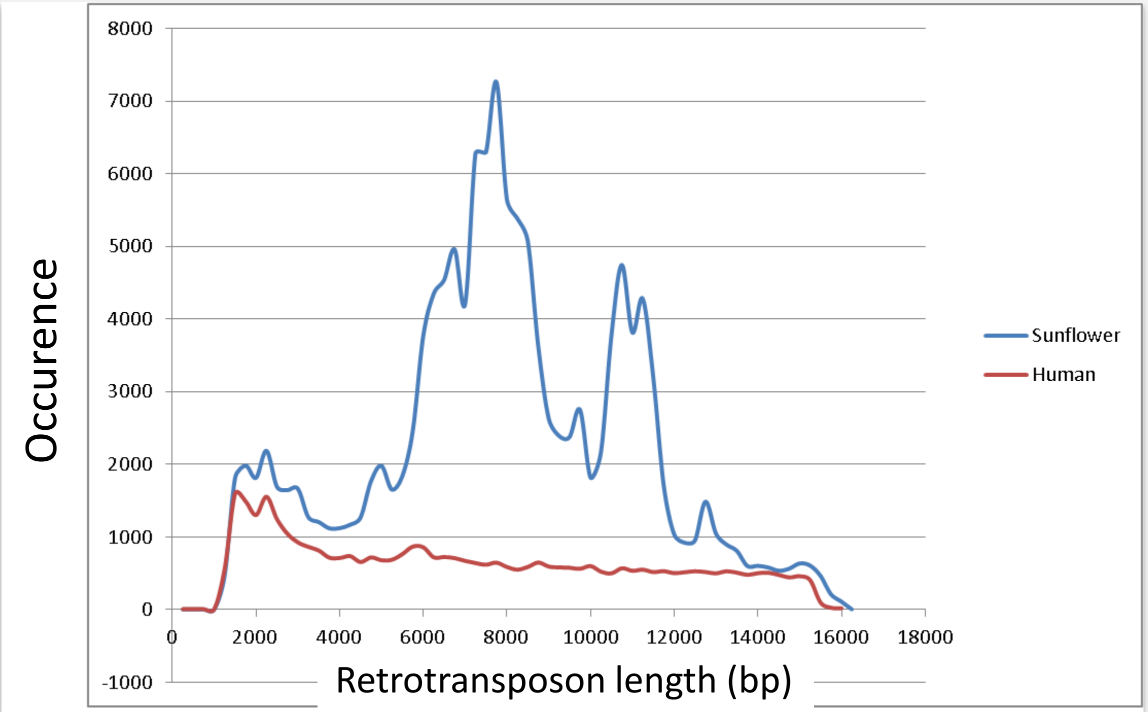
- **~ 80% known repeated sequences**
- **Majority of LTR < 0.5My**


The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements

S. Evan Staton[1], Bradley H. Bakken[2], Benjamin K. Blackman[3,†], Mark A. Chapman[4,‡], Nolan C. Kane[5], Shunxue Tang[6,§], Mark C. Ungerer[2], Steven J. Knapp[6,¶], Loren H. Rieseberg[5] and John M. Burke[4,*]

Sunflower genome contains long repeated elements

Length distribution of LTR retrotransposons



33% of sunflower genome

8% human genome

LTR up to 12kb

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

2008-2016: International Consortium for Sunflower Genomics

- Academic partners

  UBC Vancouver, INRA Toulouse, UGA Athens

- Private Partners

2015: PacBio sunflower genome

SUNRISE Project (2012-2019)

INRA Toulouse (LIPM, CNRGV, Genomics Platform)

- 407 SMRT Cells with P6/C4

-102X ➔ 4.7Tb raw data

-Acquisition of PacBio RSII at INRA Toulouse

-April – July 2015

- Novel assembly sofware

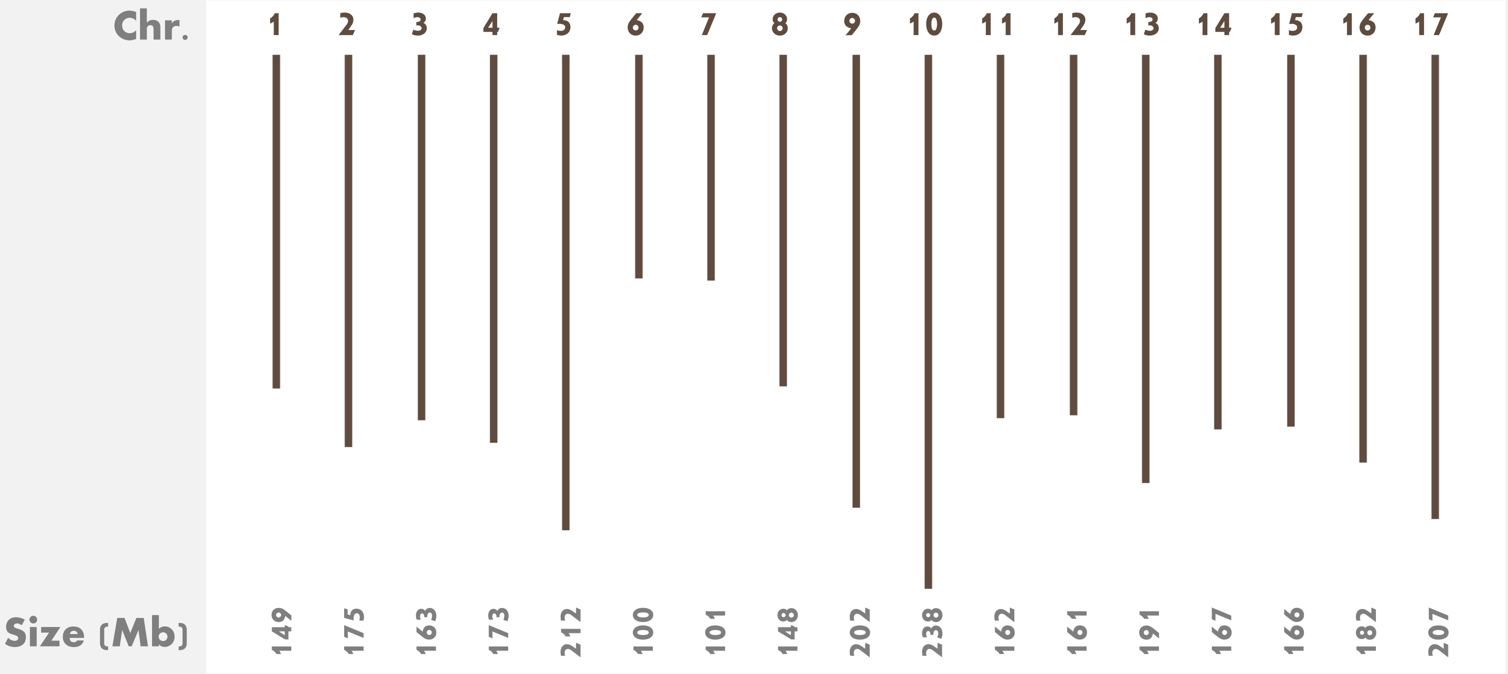
-Novel librairy preparation protocols

J. Gouzy

B. Mayjonade

Sunflower reference genome

3 027 Mb (3.38 % de N)



Mitochondrion

301 kb

Chloroplast

151 kb

80% of the genome

98.5% of contigs in the chromosomes

Gene content and genome annotation

98% of transcripts mapped on pseudo-molecules

61 RNA-Seq librairies on the sequenced genotype (XRQ)

- **Organ-specific** expression (12 organs)
- **Abiotic stress** response: drought, osmotic stress, salt stress (in roots and leaves)
- **Hormone regulation**: (9 hormones in roots and leaves)

Gene annotation

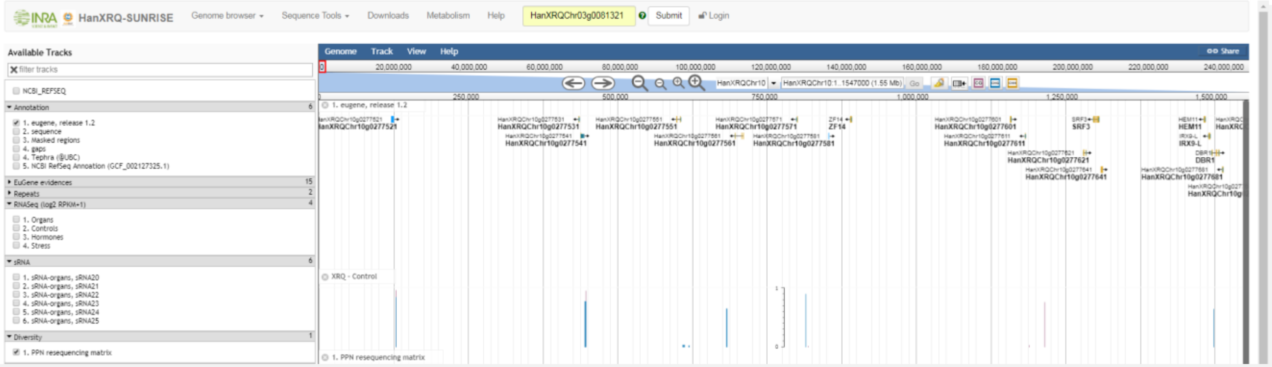
- **52 243** protein coding genes (mRNA)
- **4 945** lncRNA genes
- **88** pre-miRNA genes (351 mature miRNA)
- **862** tRNA and rRNA genes

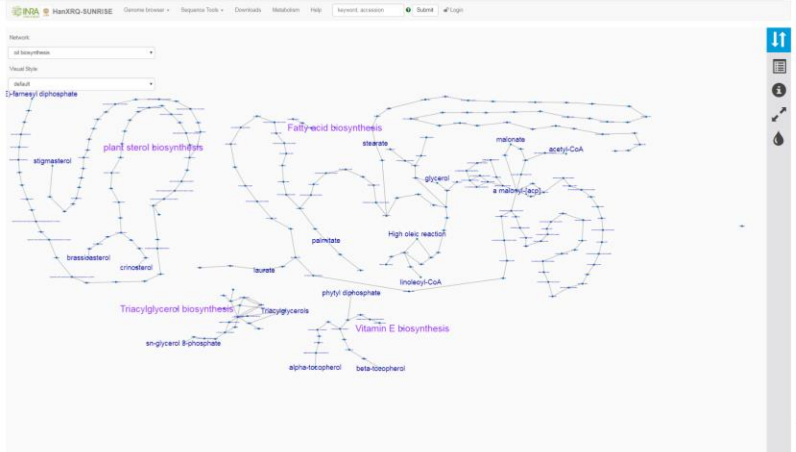
Functional genomics

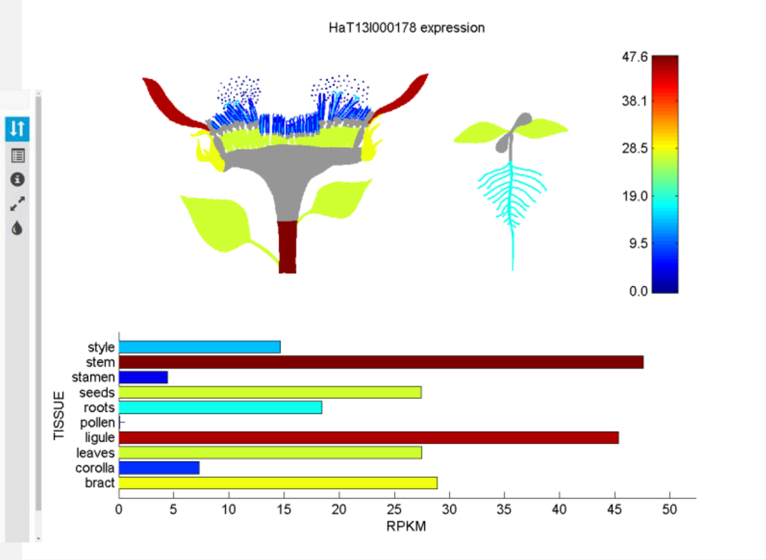
User interface and data visualization
www.heliagene.org

- Maintain access to raw data

- Integrate visualization of
  - genomic methylation
  - gene expression
  - metabolic pathways







Sequencing sunflower diversity

Cultivated sunflowers

SUNRISE: 57 public +15 private
Intl Consortium: 280 public lines

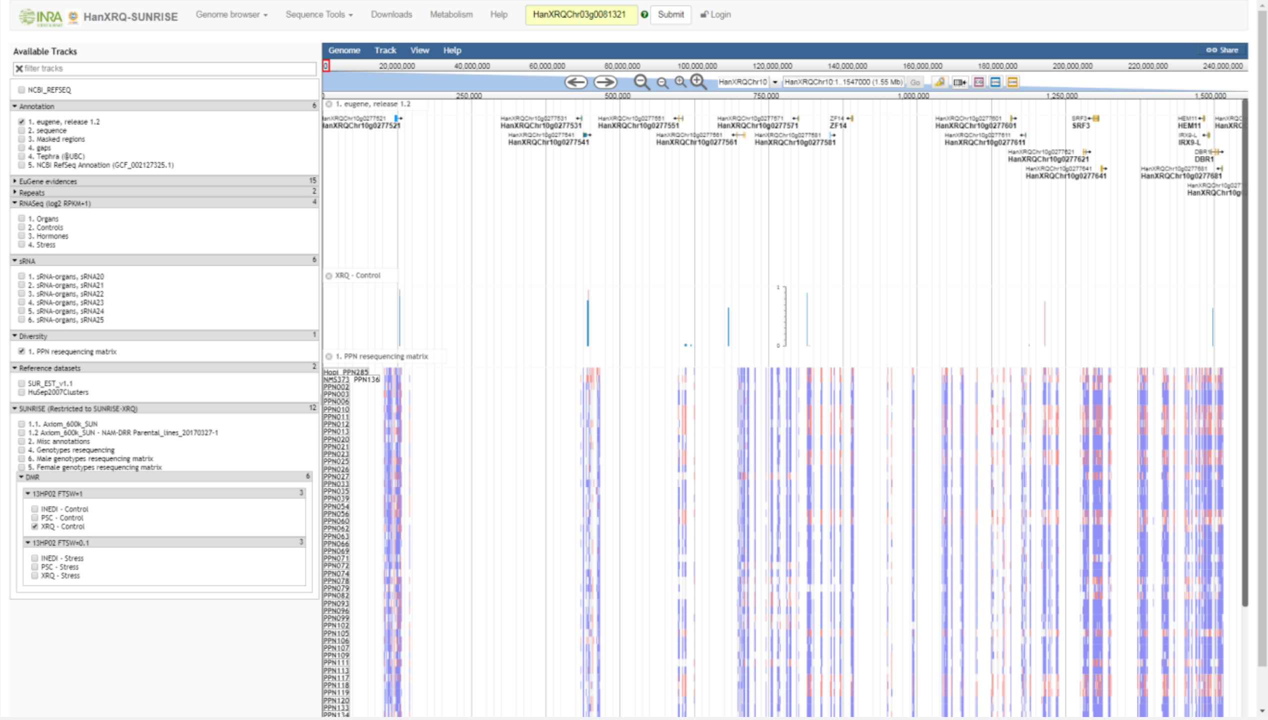
➔ RE-SEQUENCING

Sequencing sunflower diversity

Cultivated sunflowers

SUNRISE: 57 public +15 private
Intl Consortium: 280 public lines

➔ RE-SEQUENCING



Sequencing sunflower diversity

Cultivated sunflowers

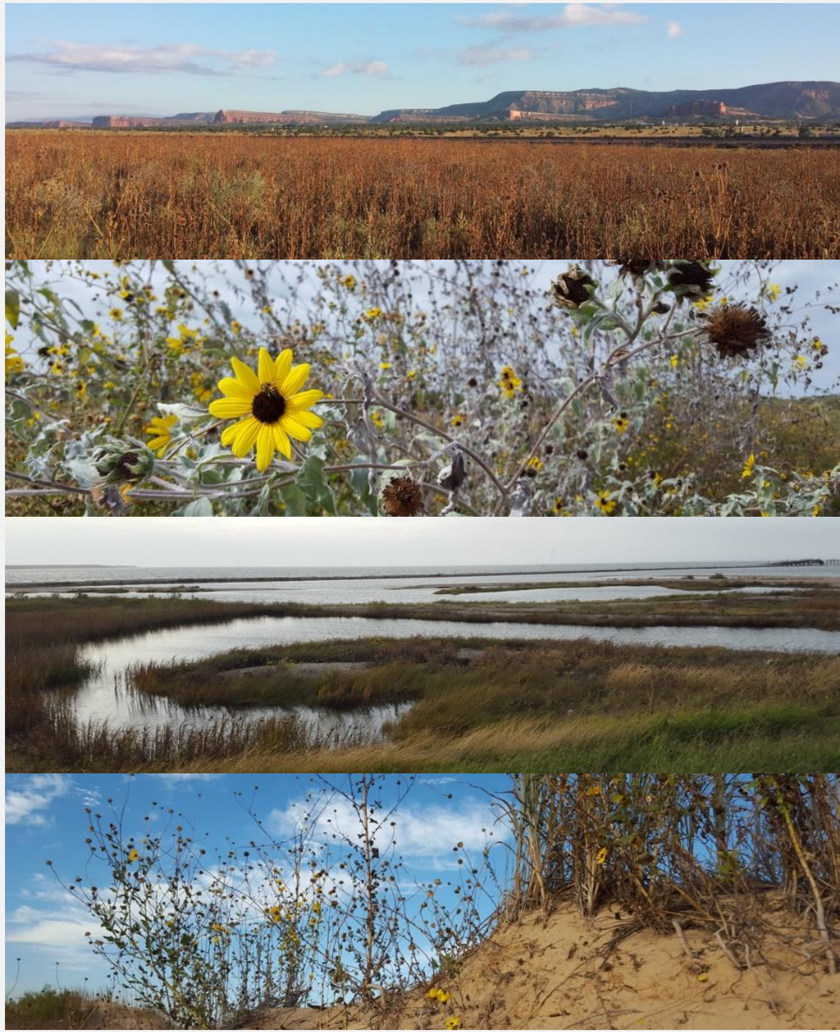
SUNRISE: 57 public +15 private
Intl Consortium: 280 public lines

➔ RE-SEQUENCING

Wild sunflowers

Intl Consortium: 3 species
H. annuus, H. argophyllus, H. petiolaris

➔ REFERENCE GENOMES *on progress*



New references: new issues

New references

In 2018/9

- New reference seq for INRA maintainer XRQ
- New high quality seq for INRA restorer line PSC8
- New high quality seq for US maintainer line HA412-HO

In 2019/20

- 3 new H. annuus H argophyllus and H. petiolaris
- 3 high quality cultivated lines

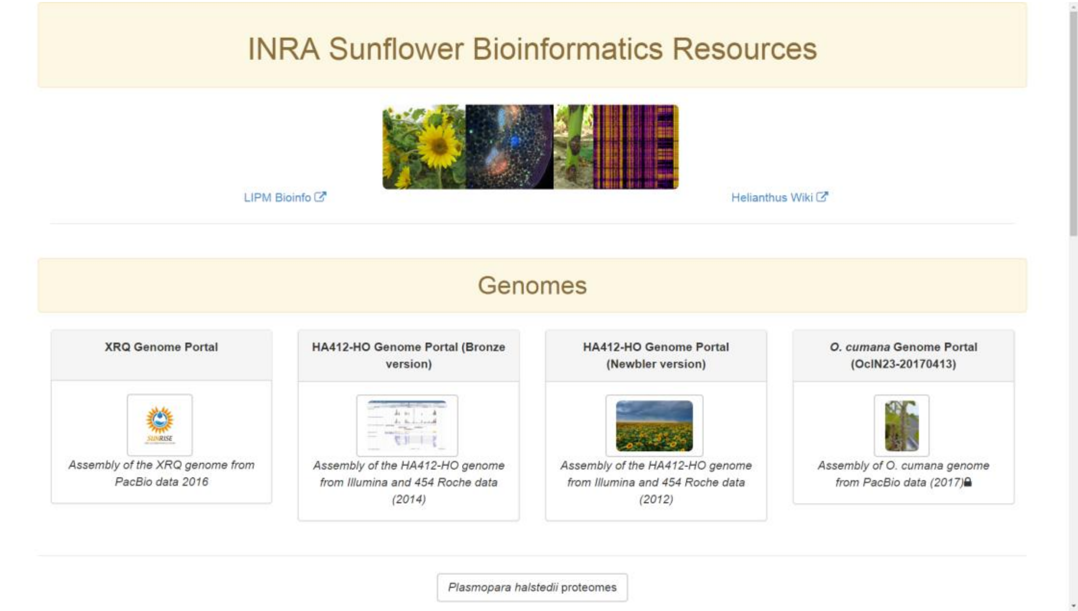
New Issues:

Computing time and expertise

- Annotation
- Genetic studies: mapping re-sequencing data (>100Tb compressed data)
- Functional genomics (expression, methylation data)

Obsolence

- Maintain web services with for the 1st reference
- Keep relations between old data / results and newest gene references



INRA Sunflower Bioinformatics Resources

LIPM Bioinfo

Helianthus Wiki

Genomes

XRQ Genome Portal

Assembly of the XRQ genome from PacBio data 2016

HA412-HO Genome Portal (Bronze version)

Assembly of the HA412-HO genome from Illumina and 454 Roche data (2014)

HA412-HO Genome Portal (Newbler version)

Assembly of the HA412-HO genome from Illumina and 454 Roche data (2012)

***O. cumana* Genome Portal (OcIN23-20170413)**

Assembly of O. cumana genome from PacBio data (2017)

Plasmopara halstedii proteomes

Genomics data: only one side of the coin

French Biological Resource Centre

Patrimonial resources:

- 2313 cultivated lines
- 403 ancient cultivated populations
- 537 wild accessions from *Helianthus* genus

Research Genetic Resources:

- >2800 lines

High Throughput Phenotyping Platforms

At the same time: sequencing AND phenotyping all our resources

Merci pour votre attention

www.sunrise-project.fr

Financeurs

Partenaires