



**HAL**  
open science

## Genome Bioinformatics at GDEC

Frédéric Choulet

► **To cite this version:**

Frédéric Choulet. Genome Bioinformatics at GDEC. AuBi-NGS Annual Meeting, AuBi - Auvergne Bioinformatique, Dec 2022, Clermont-Ferrand, France. hal-04090099

**HAL Id: hal-04090099**

**<https://hal.inrae.fr/hal-04090099>**

Submitted on 5 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

# Genome Bioinformatics at GDEC

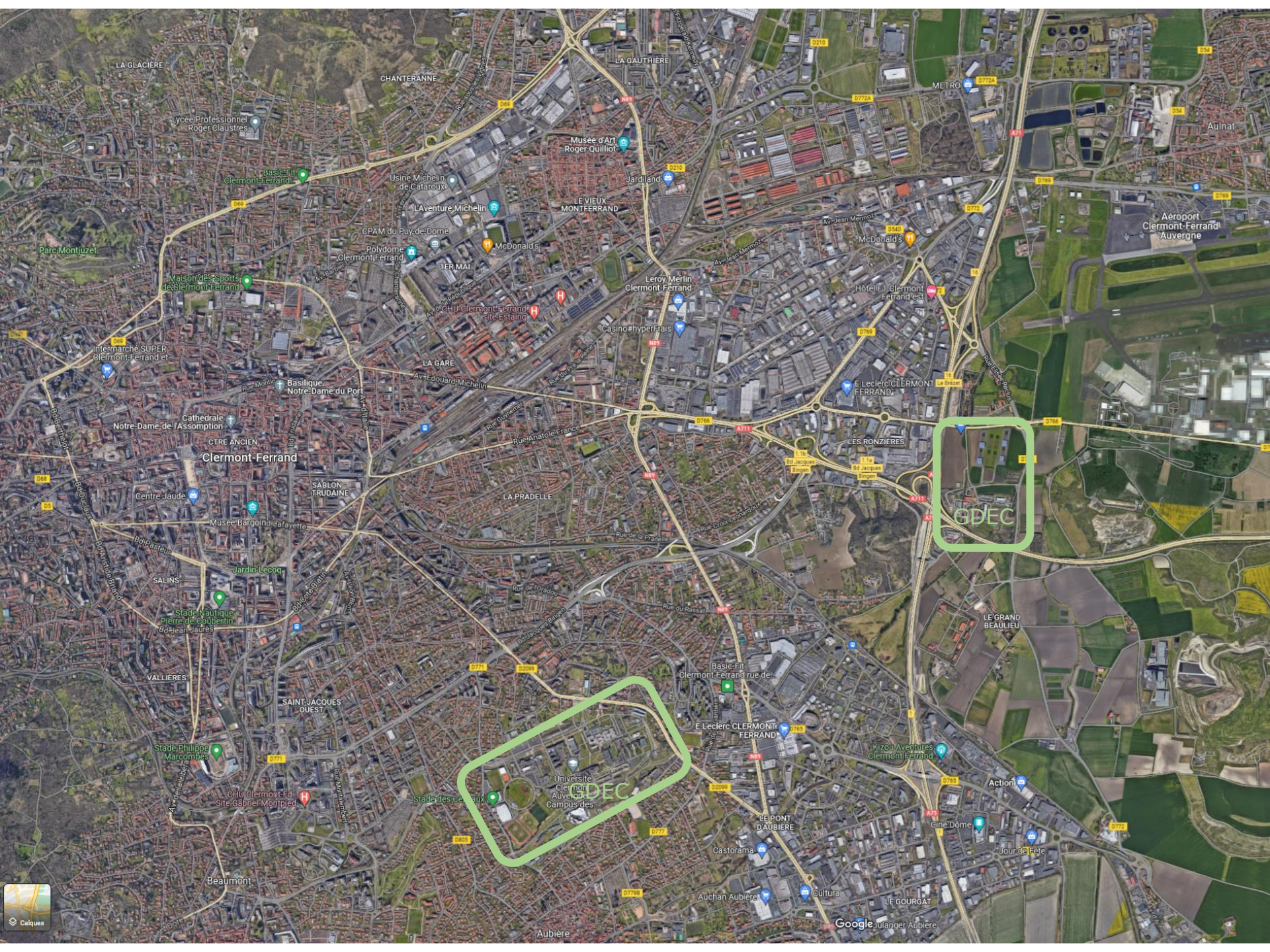
*Frédéric Choulet*



**INRAE**

**UCA**  
UNIVERSITÉ  
Clermont  
Auvergne



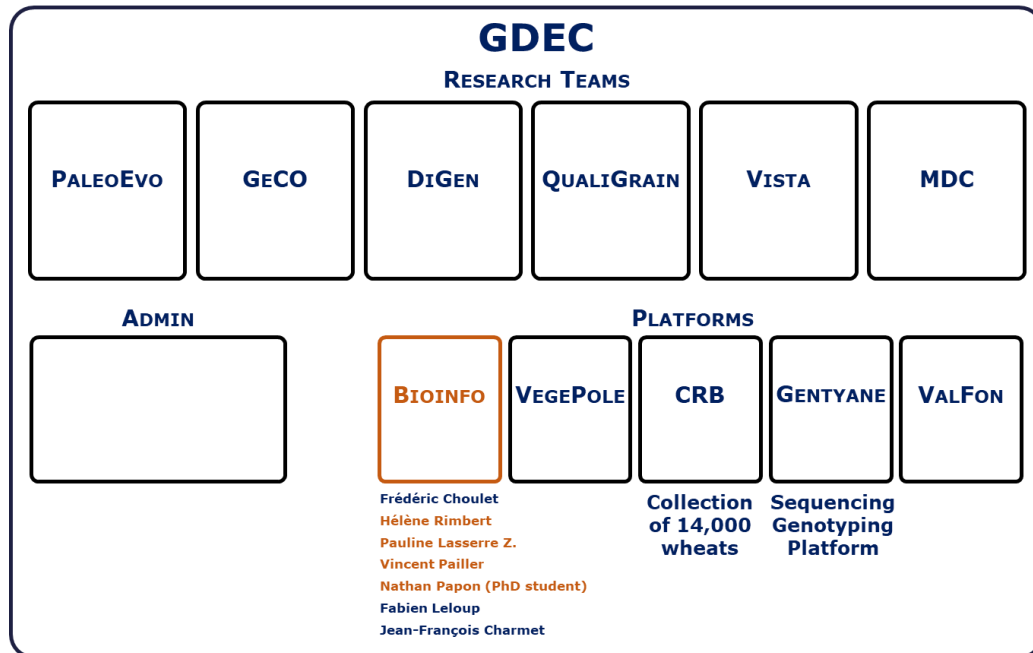


GDEC

DEC







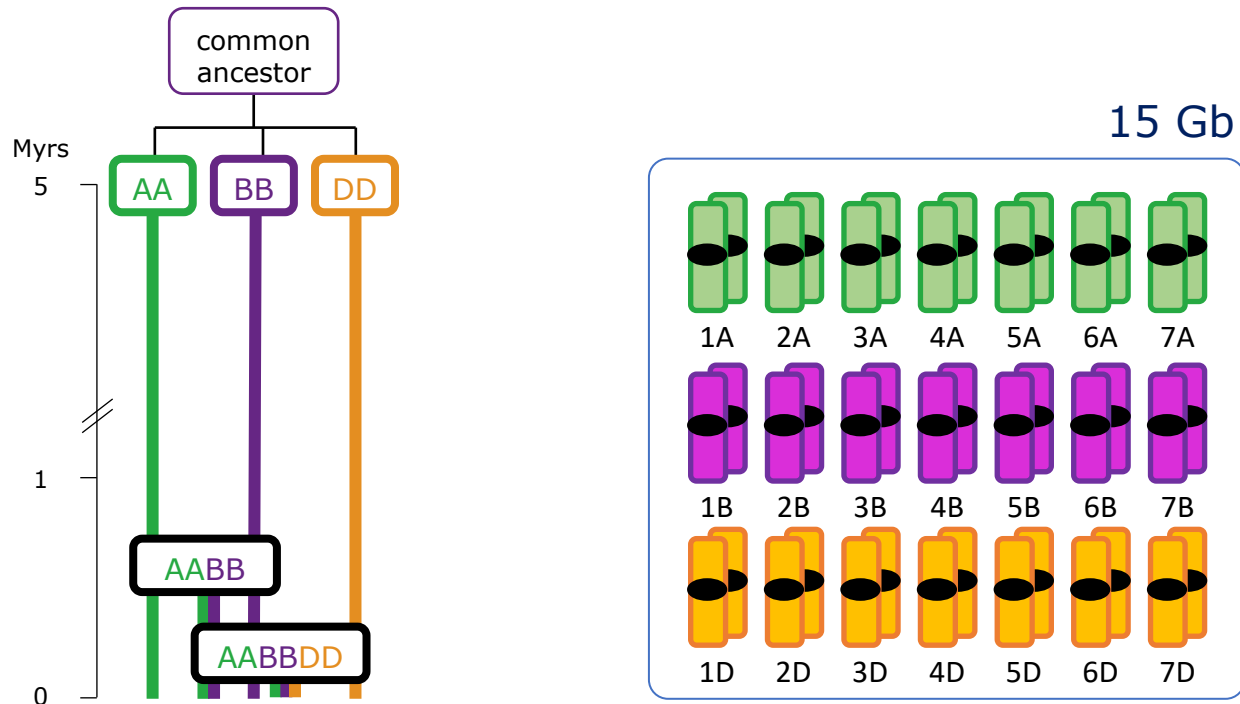
- Bioinfo for bread wheat and *Triticeae* (pan)genomics



# ❖ Research area and questions

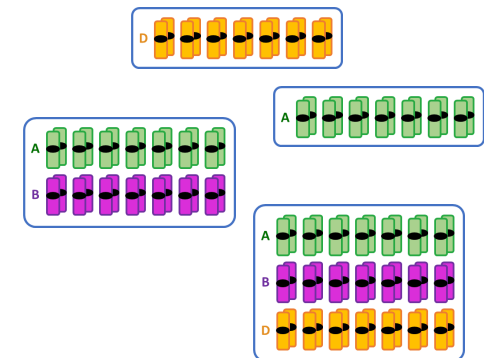
- Methodologies
  - Gene and transposon prediction in genomes
  - Comparative genomics
  - How to compare “complex” genomes?
- Genome Biology
  - What is the extent of diversity? Pangenomics
  - What is the dynamics of transposons?
  - What is the impact of polyploidy on transposon activity?
  - ...

# ❖ Hexaploid genome



## *Triticum/Aegilops*

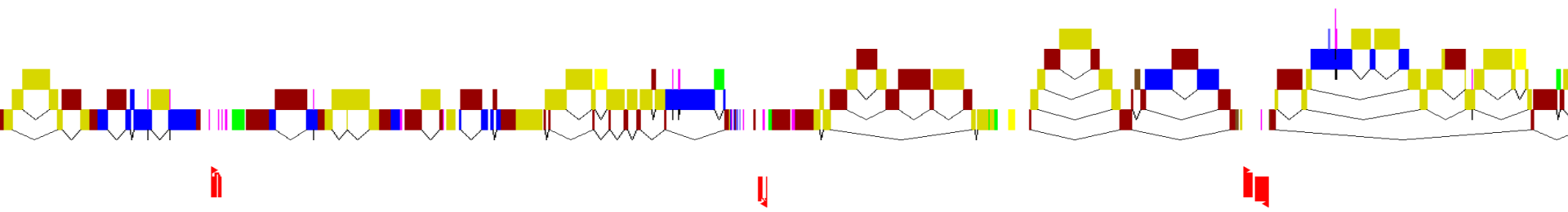
- Wild **diploids** (AA, BB, DD, SS, UU, NN, ...)
- Wild & cultivated **tetraploids** (AABB, DDNN, ...)
- Cultivated **hexaploids** (AABBDD)





# ❖ Genome composition

- Transposons (85%)
- Genes (2%)



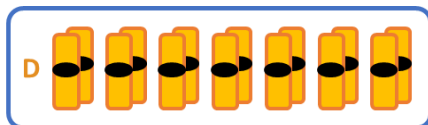
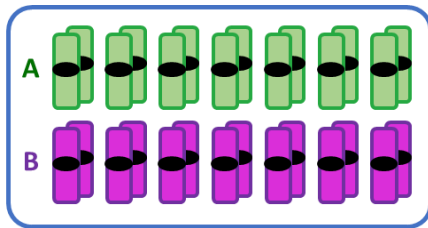
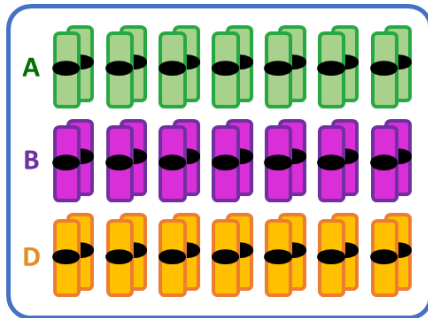
1<sup>st</sup> genome seq assembled:

**Chinese Spring** (*IWGSC Science 2018*)

- 4 million TEs
- ~107,000 genes



# ❖ Pangenomics



- *T. aestivum*

- **Chinese Spring** (Illumina)
- 10 other varieties (Illumina)
- Zang1817 (Illumina)
- Fielder (Pacbio)
- Kariega (Pacbio)
- Renan (ONT)
- Attraktion (Pacbio)

*IWGSC 2018*

*Walkowiak et al. 2020*

*Guo et al. 2020*

*Sato et al. 2021*

*Athiyannan et al. 2022*

*Aury et al. 2022*

*Kale et al. 2022*

- *T. durum* AABB

- *T. dicoccoides* AABB

- *T. urartu* AA

- *Ae. tauschii* DD

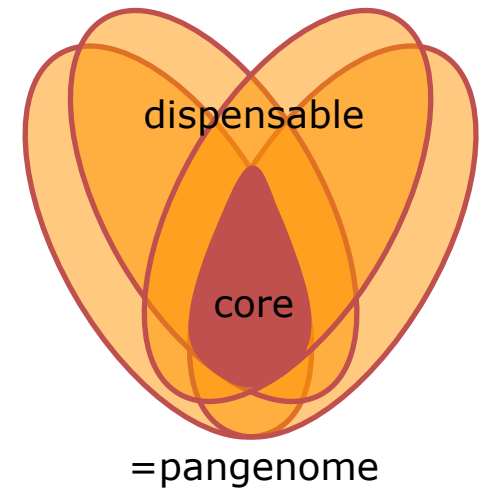
- *Ae. searsii* DD

- *Ae. bicornis* DD

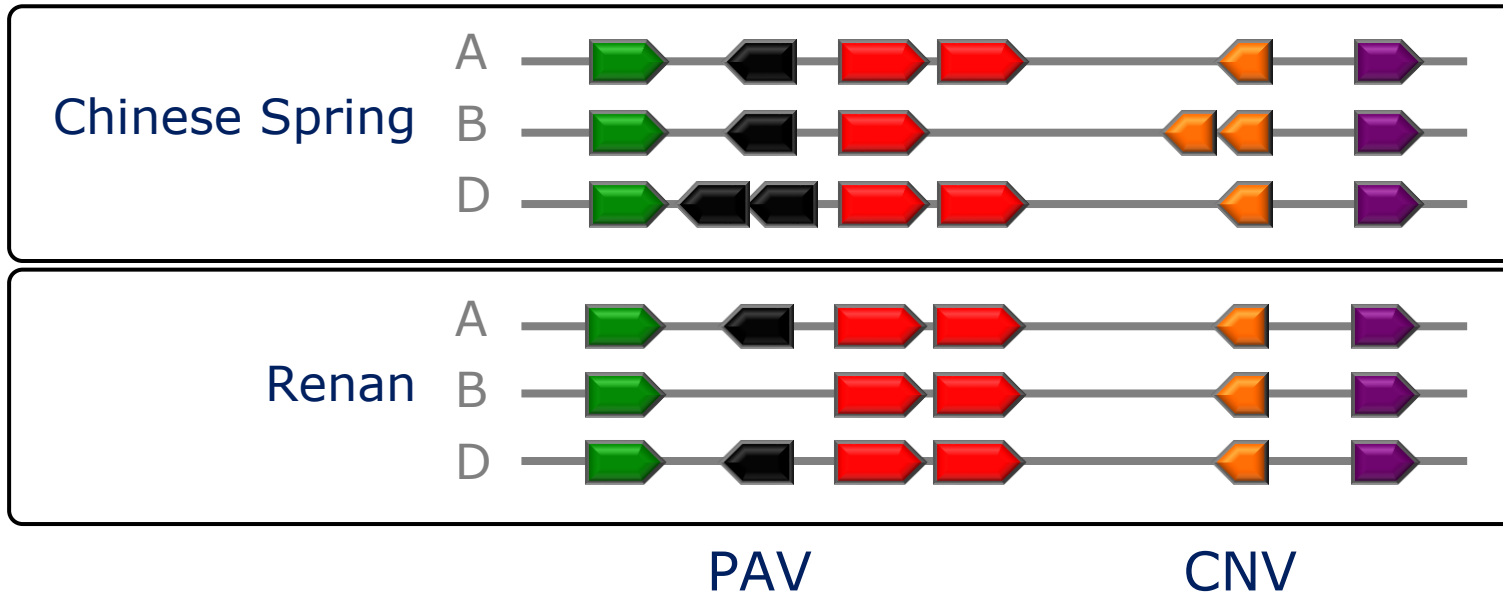
- *Ae. longissima* DD

- *Ae. sharonensis* DD

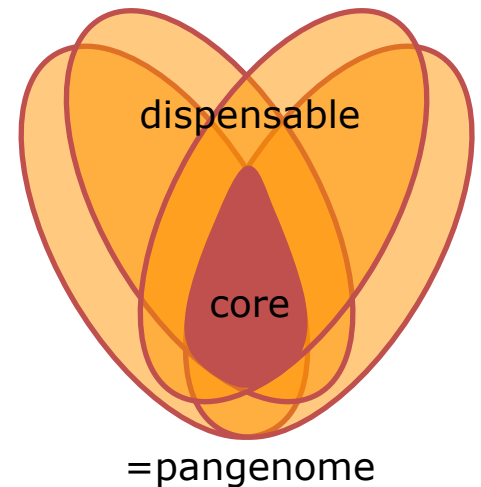
- *Ae. speltoides* SS

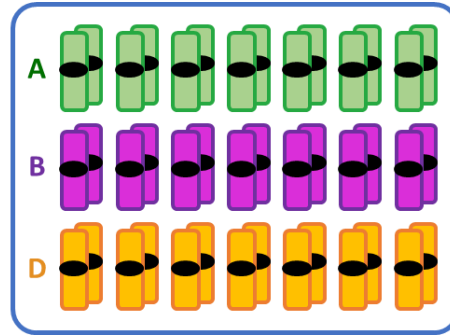


# ❖ Pangenomics

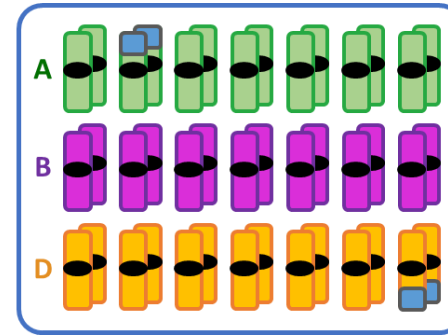


- Chromosome-scale assemblies required
- Quality of annotation matters



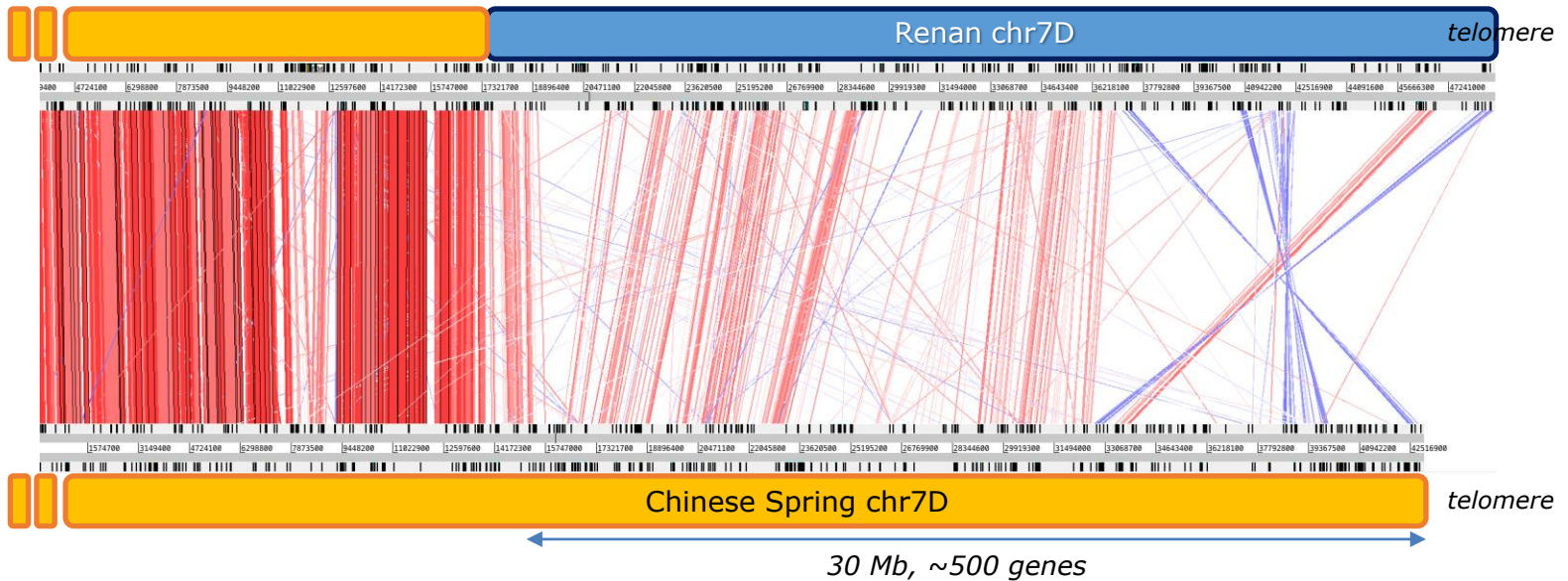


Chinese Spring



Renan

- ✓ Nanopore
- ✓ Illumina
- ✓ Bionano
- ✓ HiC

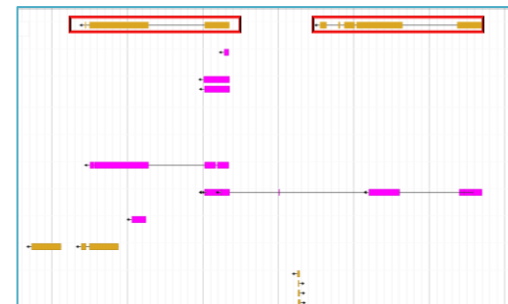




❖ **Genes**

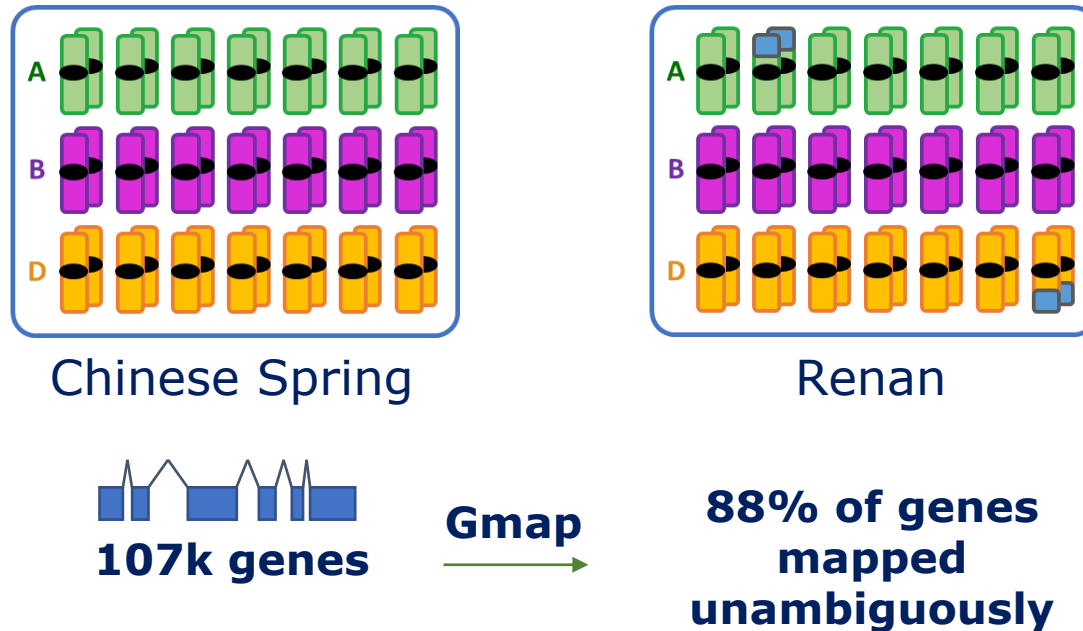
# ❖ Gene annotation

- Predicting genes is not trivial
  - *de novo* gene prediction not adapted to pangenomics



# ❖ Gene annotation

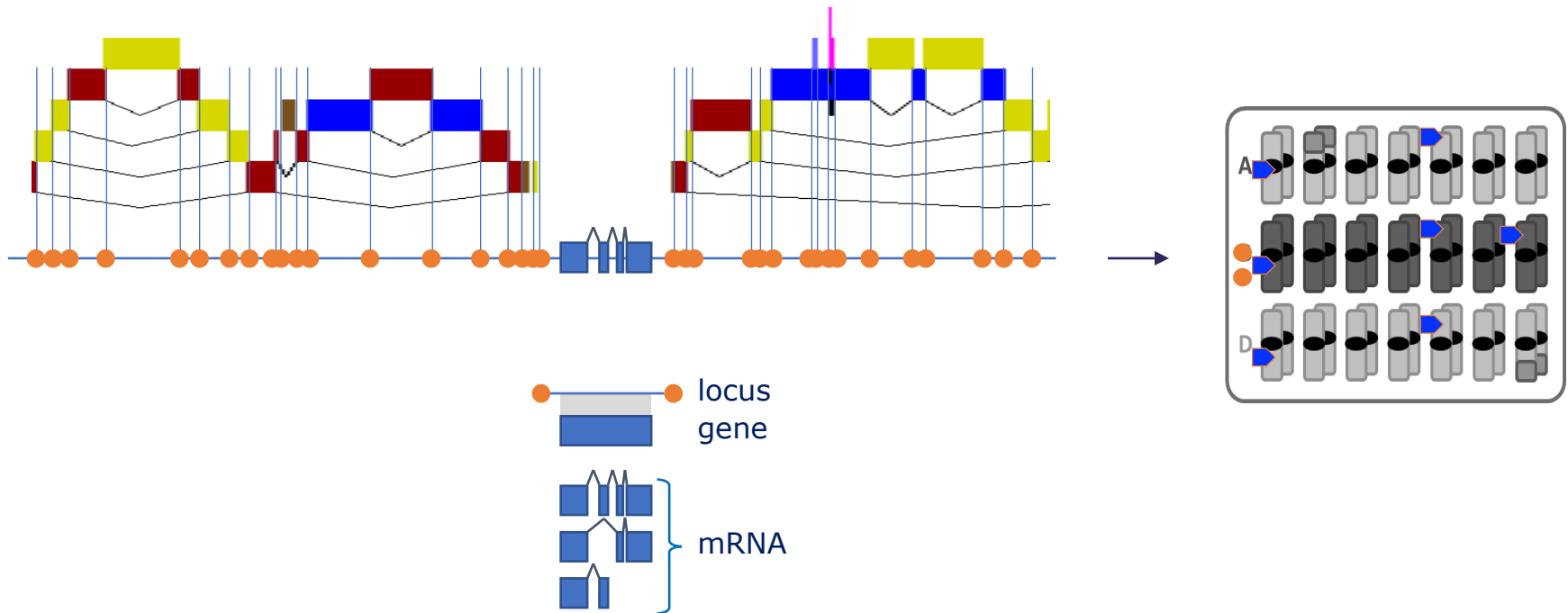
- Predicting genes is not trivial
  - *de novo* gene prediction not adapted to pangenomics
- Mapping a reference gene set sounds better
  - Develop **MAGATT** (*H. Rimbart*)





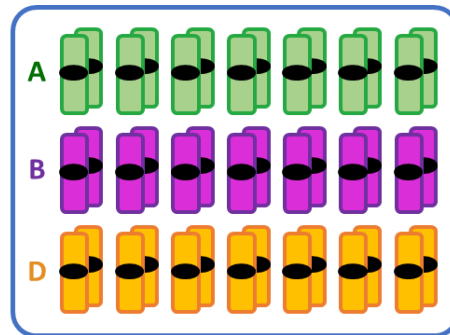
# ❖ Gene annotation

- Predicting genes is not trivial
  - *de novo* gene prediction not adapted to pangenomics
- Mapping a reference gene set sounds better
  - Develop **MAGATT** (*H. Rimbart*)

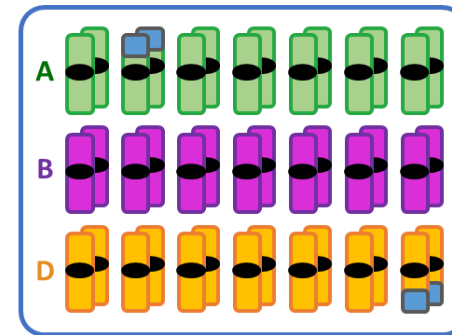


# ❖ Gene annotation

- Predicting genes is not trivial
  - *de novo* gene prediction not adapted to pangenomics
- Mapping a reference gene set sounds better
  - Develop **MAGATT** (*H. Rimbart*)



Chinese Spring



Renan

  
107k genes

Gmap  
→

88% of genes  
mapped  
unambiguously

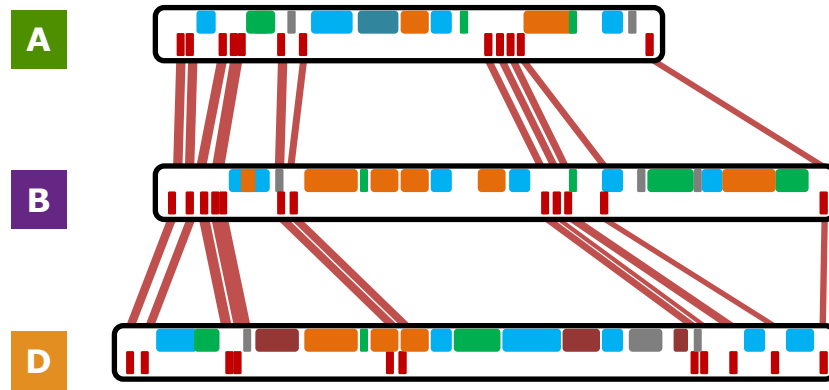
**MAGATT**  
→

**98% of genes  
mapped  
unambiguously**

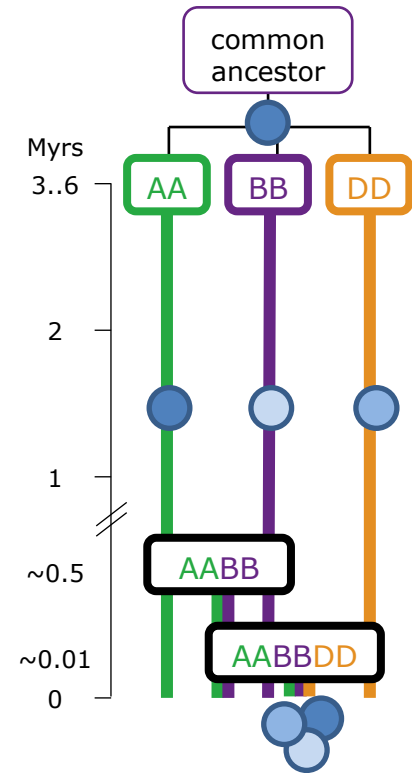
## ❖ **Transposons**



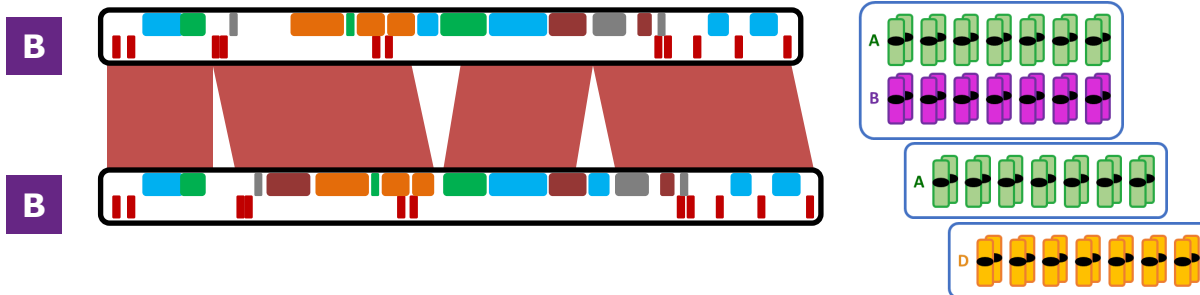
# ❖ Transposon dynamics



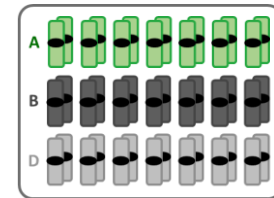
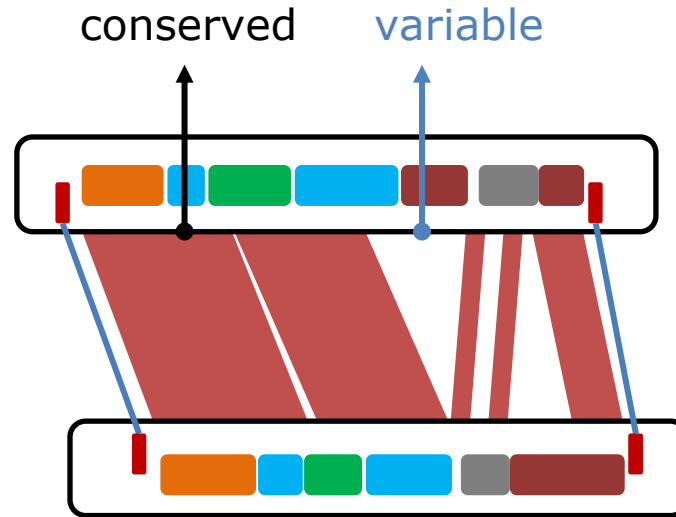
➤ **Equilibrium** model of evolution



PhD Nathan Papon (2019-2023)



# ❖ Transposon dynamics



- **Method: bp-resolved gene anchored whole-genome**

- How variable is the TE space? ➤ 5 .. 15%
- Which TE families are active? ➤ ~ all
- Impact of polyploidization? ➤ none

➤ **Transposition highly regulated; transposons may be a key role in shaping a functional genome architecture**

CSH Cold Spring Harbor Laboratory

**bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID-19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results [Follow this preprint](#)

**All families of transposable elements were active in the recent wheat genome evolution and polyploidy had no impact on their activity**

Nathan Papon, Pauline Lasserre-Zuber, H  l  ne Rimbart, Romain De Oliveira, Etienne Paux, Fr  d  ric Choulet  
doi: <https://doi.org/10.1101/2022.11.25.517938>  
This article is a preprint and has not been certified by peer review [what does this mean?]

[Open Access](#) [Full Text](#) [Info/History](#) [Metrics](#) [Preview PDF](#)

**ABSTRACT**

Bread wheat (*Triticum aestivum* L.) is a major crop and its genome is one of the largest ever assembled at reference-quality level. It is 15 Gb, hexaploid, with 85% of transposable elements (TEs). Wheat genetic diversity was mainly focused on genes and little is known about the extent of genomic variability affecting TEs, transposition rate, and the impact of polyploidy. Multiple chromosome-scale assemblies are now available for bread wheat and for its tetraploid and diploid wild relatives. In this study, we computed base pair-resolved, gene-anchored, whole genome alignments of A, B, and D lineages at different ploidy levels in order to estimate the variability that affects the TE space. We used assembled genomes of 13 *T. aestivum* cultivars (6x=AABBDD), *T. durum* (4x=AABB), *T. dicoccoides* (4x=AABB), *T. urartu* (2x=AA), and *Aegilops tauschii* (2x=DD). We show that 5 to 34% of the TE fraction is variable, depending on the species divergence. Between 400 and 13,000 novel TE insertions per subgenome were detected. We found lineage-specific insertions for nearly all TE families in di-tetra- and hexaploids. No burst of transposition was observed and polyploidization did not trigger any boost of transposition. This study challenges the prevailing idea of wheat TE dynamics and is more in agreement with an equilibrium model of evolution.

**Competing Interest Statement**

The authors have declared no competing interest.

# ❖ Acknowledgments

- *GDEC, INRAE, UCA, Clermont-Fd*

Nathan Papon  
Hélène Rimbert  
Pauline Lasserre-Zuber  
Philippe Leroy  
Romain De Oliveira  
Cécile Monat  
Jonathan Kitt  
Marion Ranoux  
Florence Exbrayat  
François Balfourier  
Etienne Paux  
Pierre Sourdille  
Sophie Bouchet

- *Collaborators*

Jean-Marc Aury's team (Genoscope)  
Arnaud Bellec (INRAE CNRGV)  
Michael Alaux (INRAE URGI)  
Thomas Wicker (U Zurich)  
Heidrun Gundlach (PGSB Munich)  
Manuel Spannagl (PGSB Munich)

- *IWGSC*

Kellye Eversole, et al.



Contact:  
[frederic.choulet@inrae.fr](mailto:frederic.choulet@inrae.fr)

- *Funding*



SRESRI/Région ARA

