



**HAL**  
open science

## Sequencing complex genomes of Triticeae

Frédéric Choulet

► **To cite this version:**

Frédéric Choulet. Sequencing complex genomes of Triticeae. Pacbio PRISM 2024, Pacbio, Apr 2024, Lisbonne, Portugal. hal-04584678

**HAL Id: hal-04584678**

**<https://hal.science/hal-04584678>**

Submitted on 23 May 2024

**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.



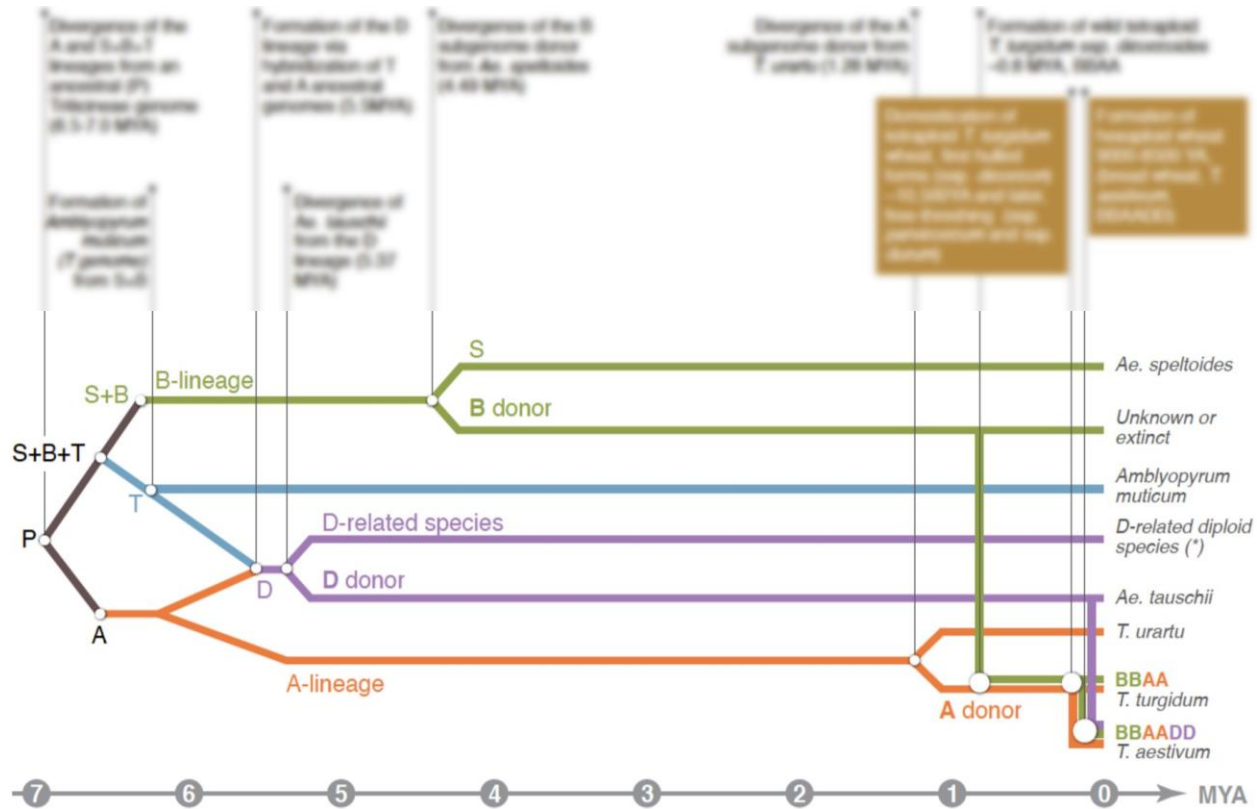
Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

# Sequencing complex genomes of *Triticeae*

PacBio PRISM 2024  
April 24<sup>th</sup> 2024 – Lisbon

Frédéric CHOLET  
[orcid.org/0000-0003-1788-7288](https://orcid.org/0000-0003-1788-7288)  
GDEC, Clermont-Ferrand, France

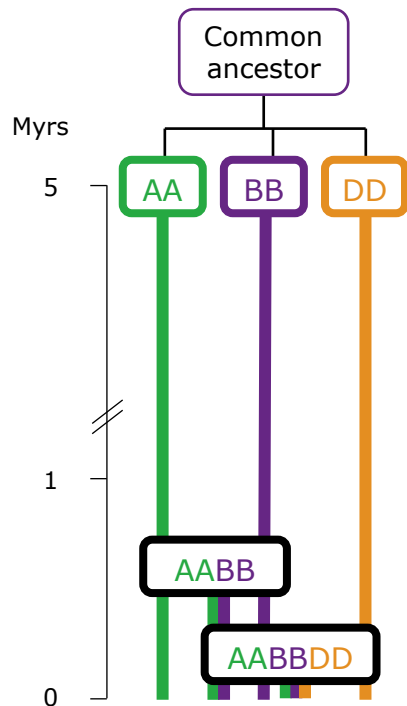
# ❖ *Triticum/Aegilops* evolution



wheat wild relatives

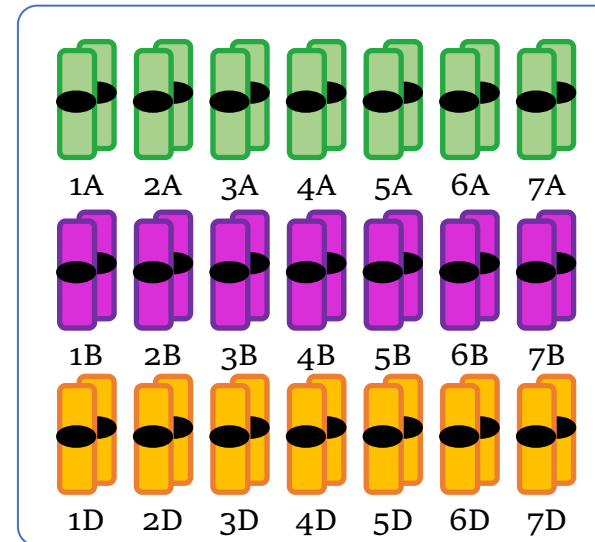
durum wheat  
bread wheat

# ❖ *Triticum/Aegilops* evolution

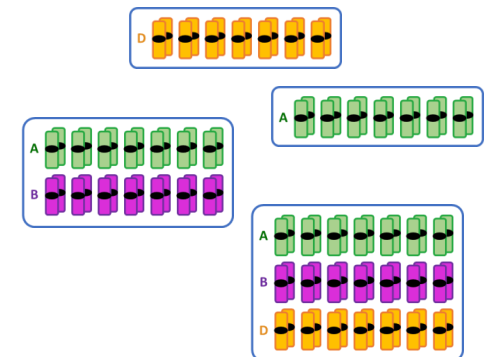


Bread wheat – *T. aestivum*

15 Gb

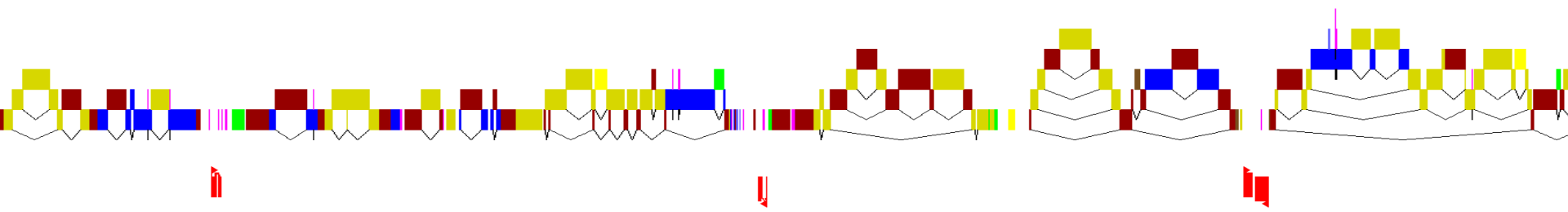


- Wild diploids (AA, BB, DD, SS, UU, NN, GG...)
- Wild and cultivated tetraploids (AABB, DDNN, ...)
- Cultivated hexaploids (AABBDD)



# ❖ *Triticeae* genome composition

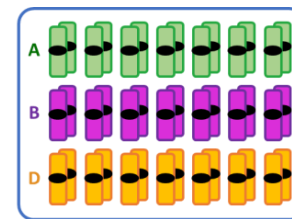
- Transposons (85%)
- Genes (2%)



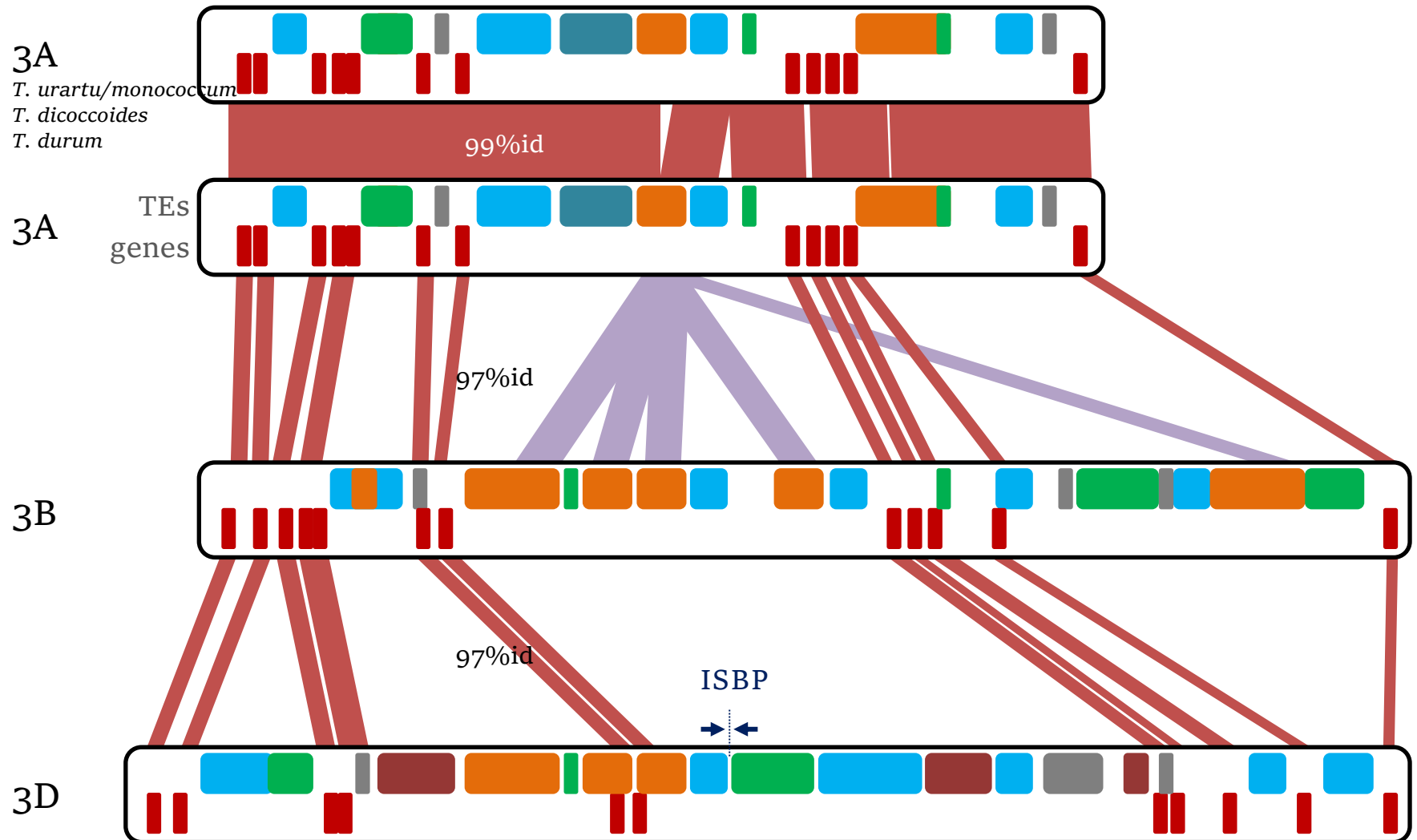
1<sup>st</sup> assembled genome:

Chinese Spring (*IWGSC Science 2018*)

- 3x ~1.3 million TEs
- 3x ~35,000 genes



# ❖ A-B-D comparative genomics



# ❖ History of sequencing *Triticeae* genomes

## ■ *T. aestivum* AABBD

|   |      |                                |                        | <u>N50</u>         |       |
|---|------|--------------------------------|------------------------|--------------------|-------|
| • | 2014 | WGS (Illumina)                 | IWGSC 2014             | 90,000 contigs/chr | 5 kb  |
|   |      | 1 chr BAC-by-BAC (Roche/454)   | Choulet et al. 2014    |                    | 1 Mb  |
| • | 2017 | WGS (Illumina+Pacbio CLR)      | Zimin et al. 2017      |                    |       |
| • | 2018 | RefSeq (NRGene Illumina)       | IWGSC 2018             |                    | 23 Mb |
| • | 2020 | 10 genotypes (NRGene Illumina) | Walkoviak et al. 2020  |                    |       |
| • | ..   | Renan (ONT+Illumina)           | Aury et al. 2022       |                    | 48 Mb |
| • | ..   | Fielder (Pacbio HiFi)          | Sato et al. 2021       | 68 contigs/chr     | 21 Mb |
| • | ..   | Kariega (Pacbio HiFi)          | Athiyannan et al. 2022 | 15 scaff/chr       | 30 Mb |
| • | ..   | Attraktion (Pacbio HiFi)       | Kale et al., 2022      | 74 contigs/chr     | 17 Mb |

## ■ Related species

*T. durum* AAB

*T. dicoccoides* AAB

*T. urartu* A

*Ae. speltoides* S

*Ae. tauschii* D

*Ae. searsii* D

*Ae. bicornis* D

*Ae. longissima* D

*Ae. sharonensis* D

*Ae. umbellulata* U

# ❖ Sequencing wheat wild relatives

Wild relatives: source of genetic diversity of major interest for future wheat improvement

- *T. timopheevii* **AAGG**  
France2030 PEPR A&N AgroDiv  
(Salse J, Pont C et al.)



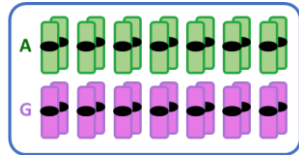
- *Ae. ventricosa* **NNDD**  
ANR 2023-26 DeFI-Wheat (P. Sourdille et al.)  
“Decipher the Fate of Introgressions from wild relatives in the wheat genome”



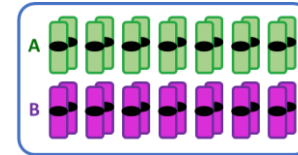
- *T. aestivum* Courtot **AABBDD**



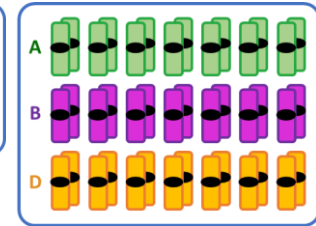
# ❖ *Triticum timopheevii*



*T. timopheevii*  
AAGG



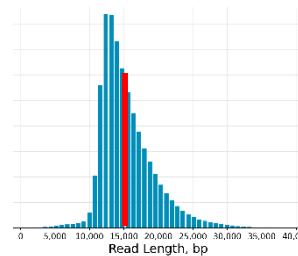
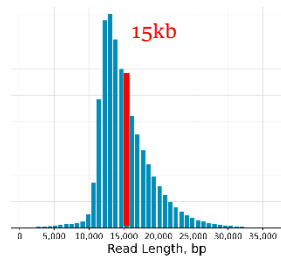
*T. durum*  
AABB



*T. aestivum*  
AABBDD

HiFi read production:  
2 smrtcells on REVIO (Gentyane)

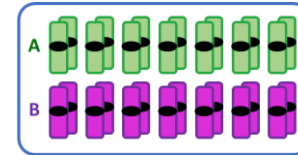
|       |       |       |       |
|-------|-------|-------|-------|
| #ccs  | 4.8 M | 5.5 M |       |
| Gbps  | 73 Gb | 83 Gb |       |
| depth | 8x    | 9x    | (17x) |
| avg   | 15 kb | 15 kb |       |



# ❖ *Triticum timopheevii*



*T. timopheevii*  
AAGG



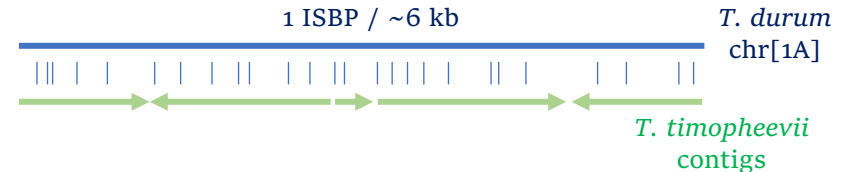
*T. durum*  
AABB

Assembly: hifiasm-0.19.8-r603

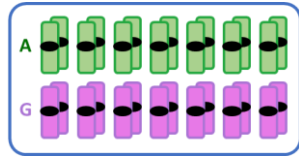
- 9.4 Gb
- 3101 contigs (max=142 Mb)
- N50=19 Mb (119 contigs)

Building 14 pseudomolecules

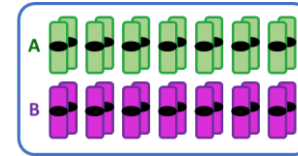
- 97% of bps anchored
  - 41 contigs / chr[1-7]A
  - 89 contigs / chr[1-7]G



# ❖ *Triticum timopheevii*

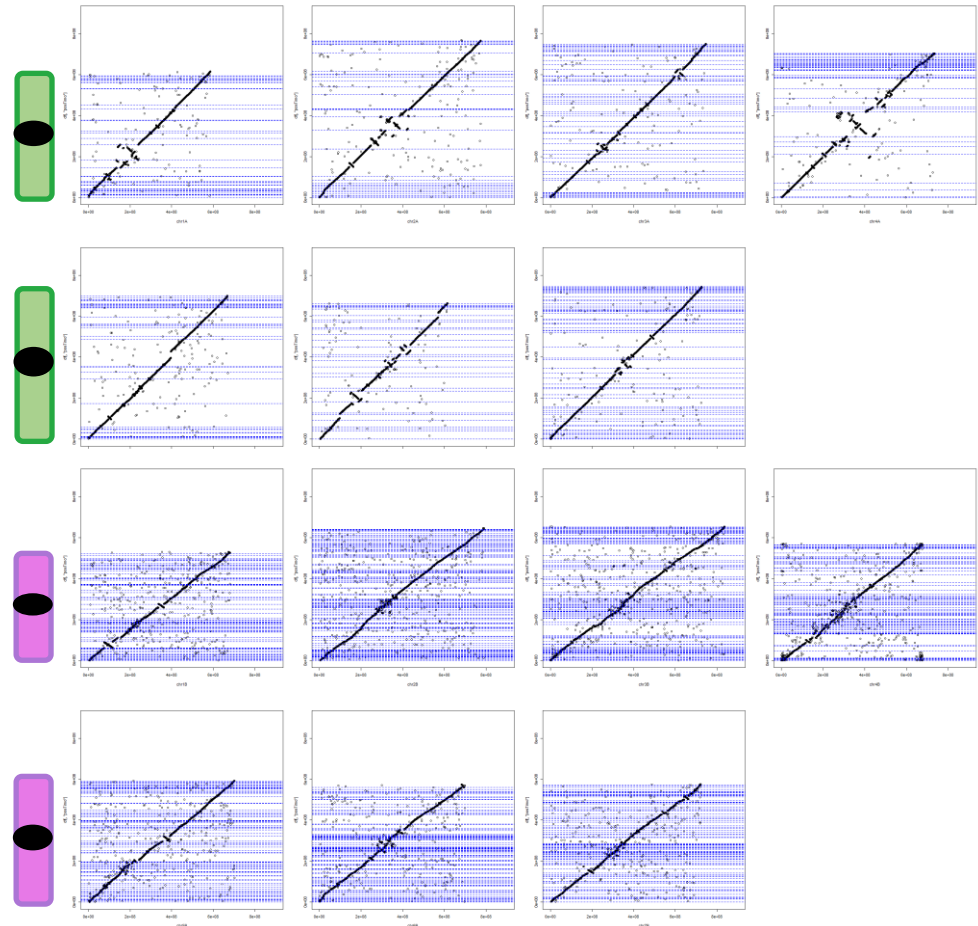


*T. timopheevii*  
AAGG



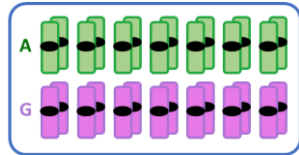
*T. durum*  
AABB

- 41 contigs / chr[1-7]A
- 89 contigs / chr[1-7]G

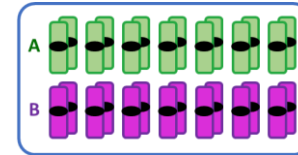




# ❖ *Triticum timopheevii*



*T. timopheevii*  
AAGG



*T. durum*  
AABB

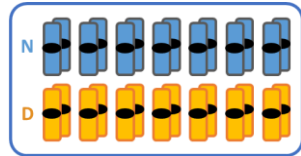
## ○ Comparative genomics:

BLAST search: 440,253 exons (*T. durum* AABB)

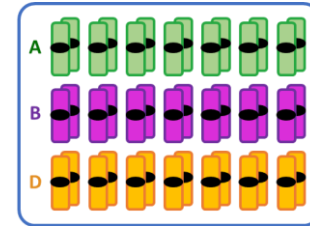
=> 98% conserved in *T. timopheevii*

- *T. durum* AA vs. *T. timopheevii* AA: 99.0%id
- *T. durum* BB vs. *T. timopheevii* GG: 98.5%id

# ❖ *Aegilops ventricosa*



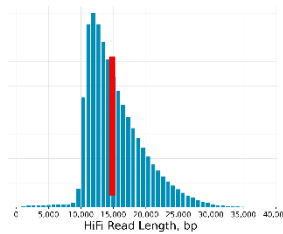
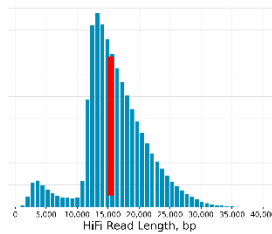
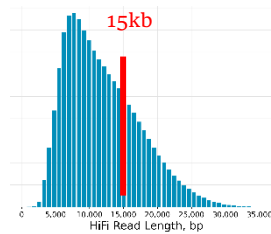
*Ae. ventricosa*  
NNDD



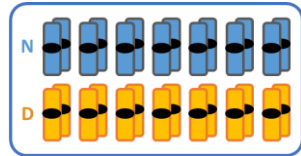
*T. aestivum*  
AABBDD

HiFi read production:  
3 smrtcells on REVIO (Gentyane)

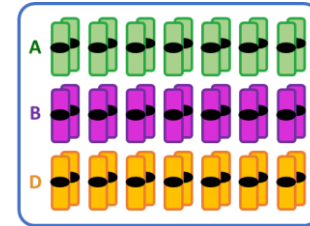
|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| #ccs  | 7.1 M | 5.8 M | 4.0 M | (23x) |
| Gbps  | 84 Gb | 94 Gb | 62 Gb |       |
| depth | 8x    | 9x    | 6x    |       |
| avg   | 12 kb | 16 kb | 15 kb |       |



# ❖ *Aegilops ventricosa*



*Ae. ventricosa*  
NNDD



*T. aestivum*  
AABBDD

Assembly: hifiasm-0.19.8-r603

- 8.6 Gb
- 2887 contigs (max=225 Mb)
- N50=39 Mb (63 contigs)

Building chr1-7D pseudomolecules

- 121 contigs ordered

15:chr1D

19:chr2D

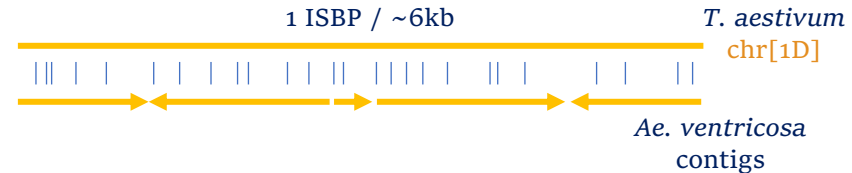
16:chr3D

17:chr4D

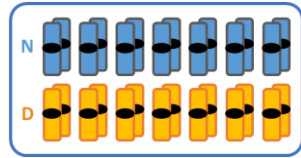
13:chr5D

24:chr6D

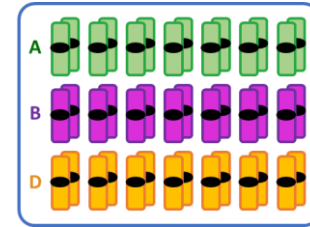
17:chr7D



# ❖ *Aegilops ventricosa*



*Ae. ventricosa*  
NNDD



*T. aestivum*  
AABBDD

4.0 Gb

- 121 contigs ordered

15:chr1D

19:chr2D

16:chr3D

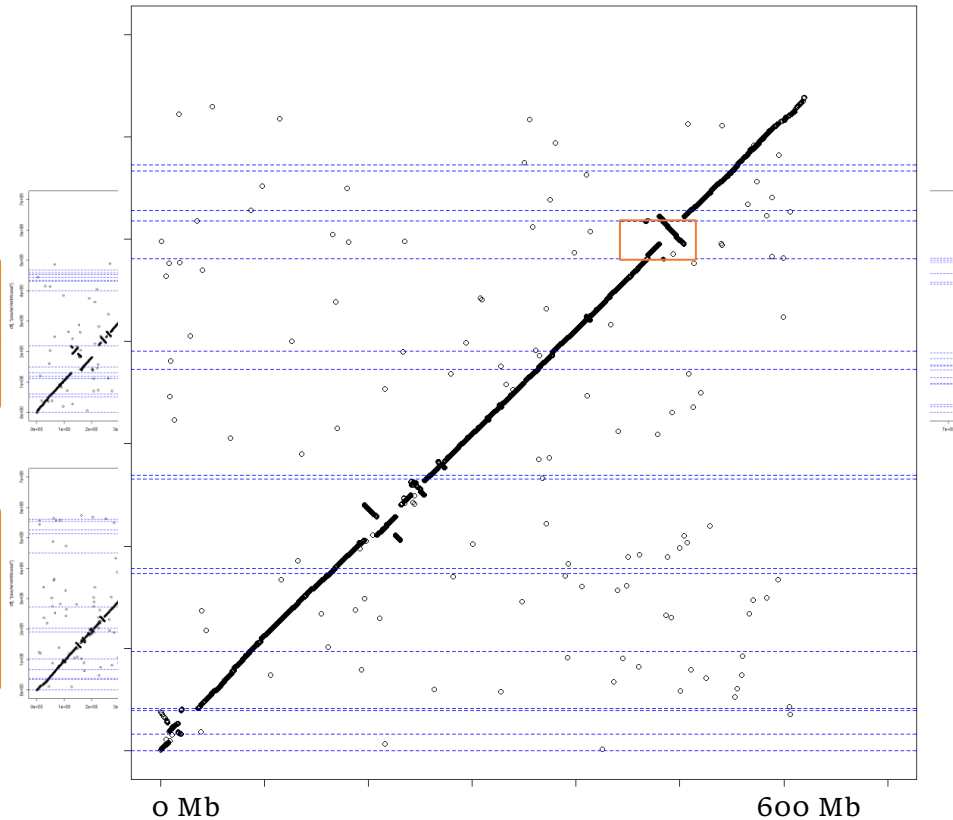
17:chr4D

13:chr5D

24:chr6D

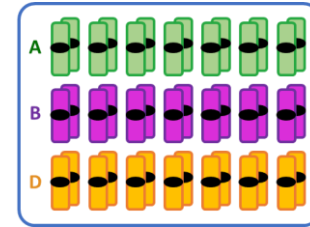
17:chr7D

4.0 Gb





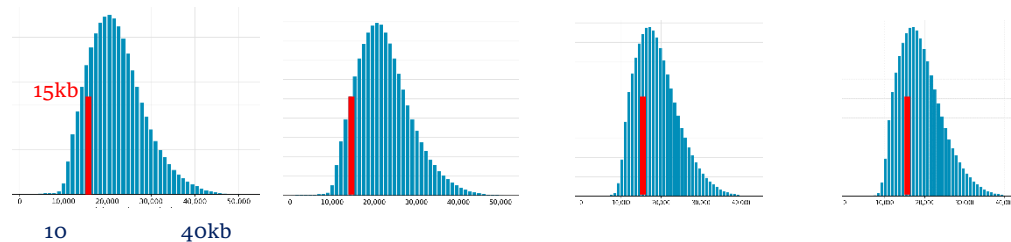
# ❖ *T. aestivum* Courtot



*T. aestivum*  
AABBDD

HiFi read production:  
4 smrtcells on REVIO (Gentyane)

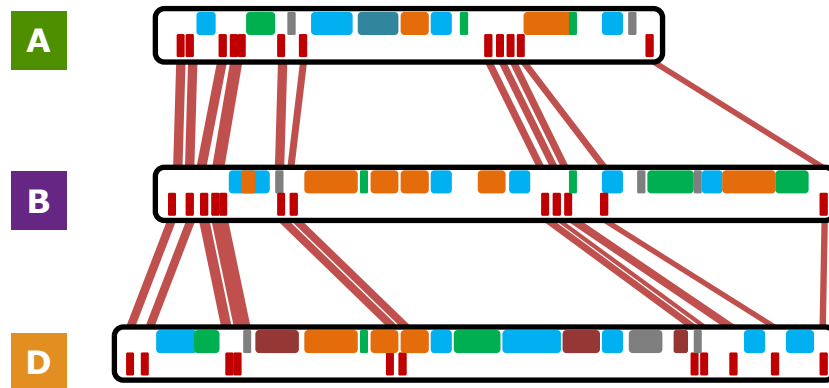
|       |       |       |        |       |       |
|-------|-------|-------|--------|-------|-------|
| #ccs  | 3.5 M | 3.3 M | 5.3 M  | 5.2 M | (24x) |
| Gbps  | 78 Gb | 73 Gb | 100 Gb | 98 Gb |       |
| depth | 5x    | 5x    | 7x     | 7x    |       |
| avg   | 22 kb | 22 kb | 19 kb  | 19 kb |       |



Assembly: hifiasm-0.19.8-r603

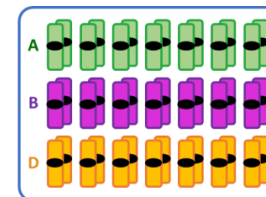
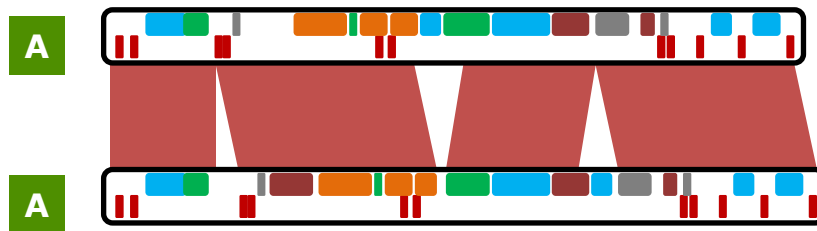
- 1744 contigs (max=**278 Mb**)
- N50=**41 Mb**

# ❖ Transposable elements

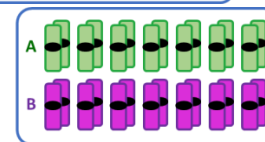


A-B-D: complete TE turnover  
TE evolution model?

- Wicker et al. *Genom Biol* 2018



*T. aestivum*



*T. durum*

*T. dicoccoides*

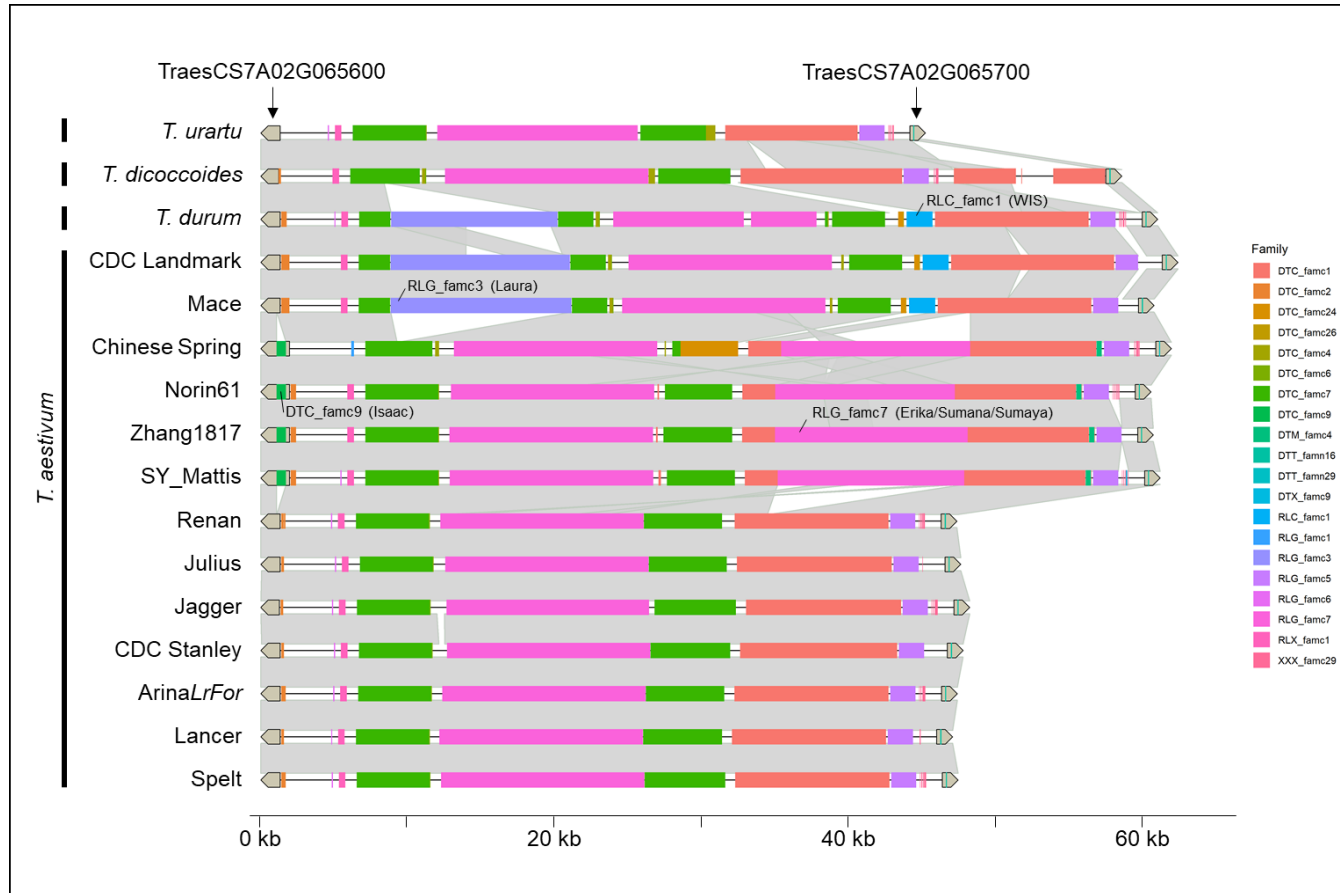


*T. urartu*

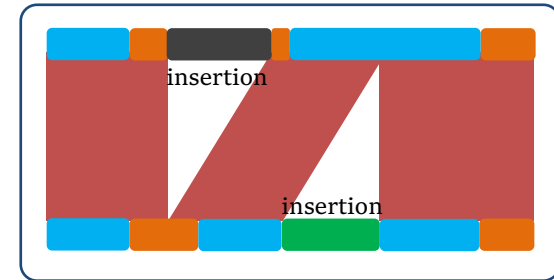
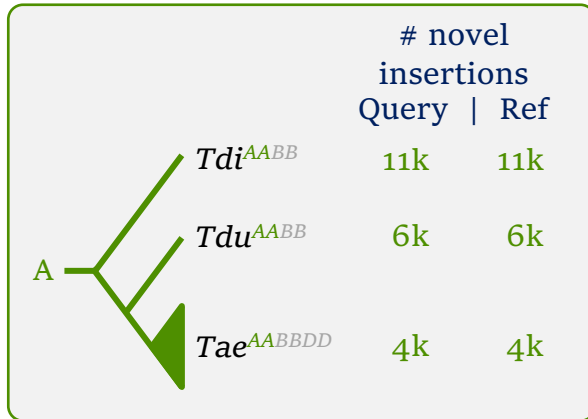


*Ae. tauschii*

# ❖ Transposable elements



# ❖ Transposable elements



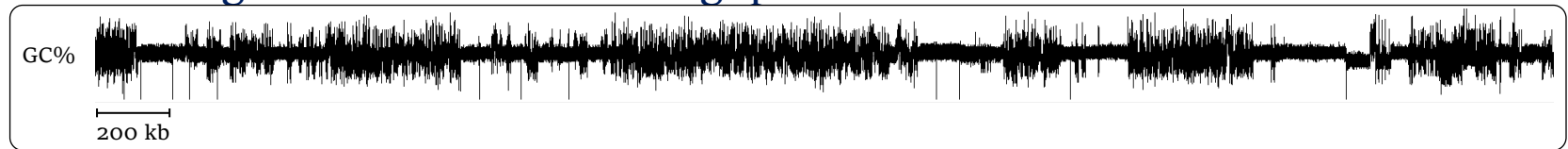
~4,000–13,000 novel insertions / subgenome (~1%)

- Transposition rate is ~ constant
  - No impact of polyploidy
  - No TE burst
  - All families produced novel copies recently
- Equilibrium model of evolution

# ❖ Conclusions

- Pacbio HiFi unlocked sequencing of *Triticeae* (complex) genomes
- Assembly stops at satellite repeats

Ultra-long reads useful for T2T gap-less assemblies



- 25x ~10..30 contigs per chr
- Bionano map still useful to reach gold standard
- Hi-C map required to build pseudomolecules

## ❑ Perspectives

wheat pangenome in progress

12 wheat landraces representing the worldwide diversity

=> starts in 2024 (Pacbio-Bionano-HiC + IsoSeq)

- ERC project “Wheat3D”
  - IPS2 (Benhamed, Latrasse)
  - URGI (Alaux)
  - INRAE GDEC

- NSF project “WheatDiversity”
  - IWGSC (J. Stein, Eversole)
  - U Georgia (Buell)
  - Corteva Agrosience (Johnston)
  - INRAE GDEC

# ❖ Acknowledgments



## ❖ GDEC INRAE-UCA

### *GENTYANE*

E. Belmonte

V. Gautier

C. Poncet

### *Team BIOINFO*

V. Pailler

H. Rimbert

P. Lasserre-Zuber

F. Leloup

### *Team GeCO*

P. Sourdille et al.

### *Team PALEOEVO*

J. Salse

C. Pont et al.

### *Team DiGEN*

S. Bouchet

M. Ranoux

J. Kitt et al.



D. Grimbichler

