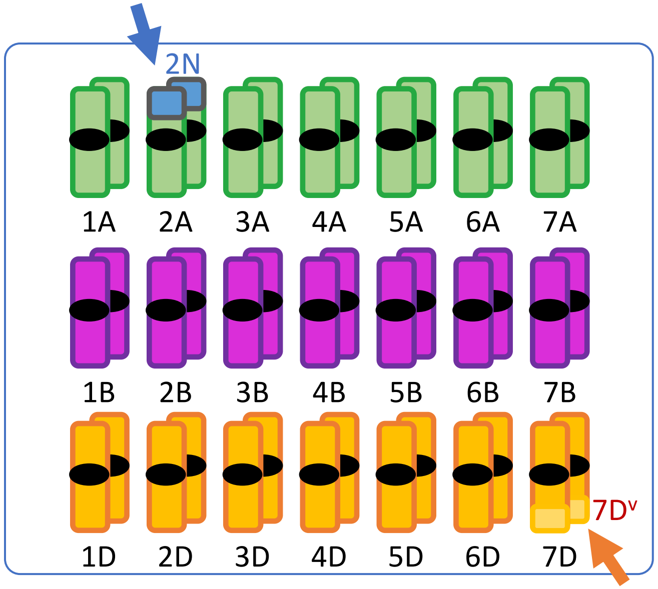
# Sequencing the genome of the French wheat variety Renan

- *Frédéric CHOULET*

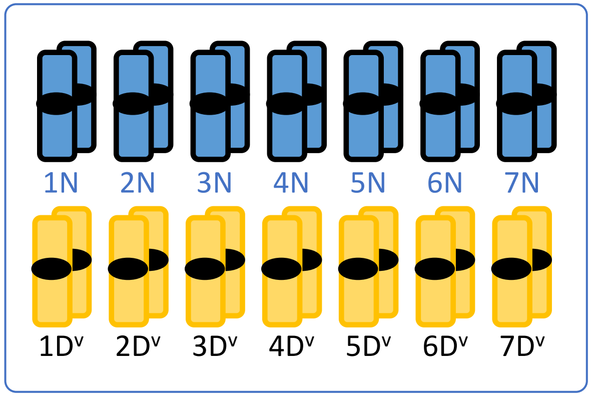
*GDEC, INRAE-Univ Clermont-Ferrand*

Yr17 Lr37 Sr38 Cre5

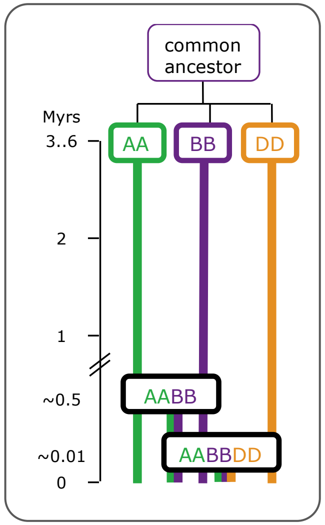


Pch1

RENAN

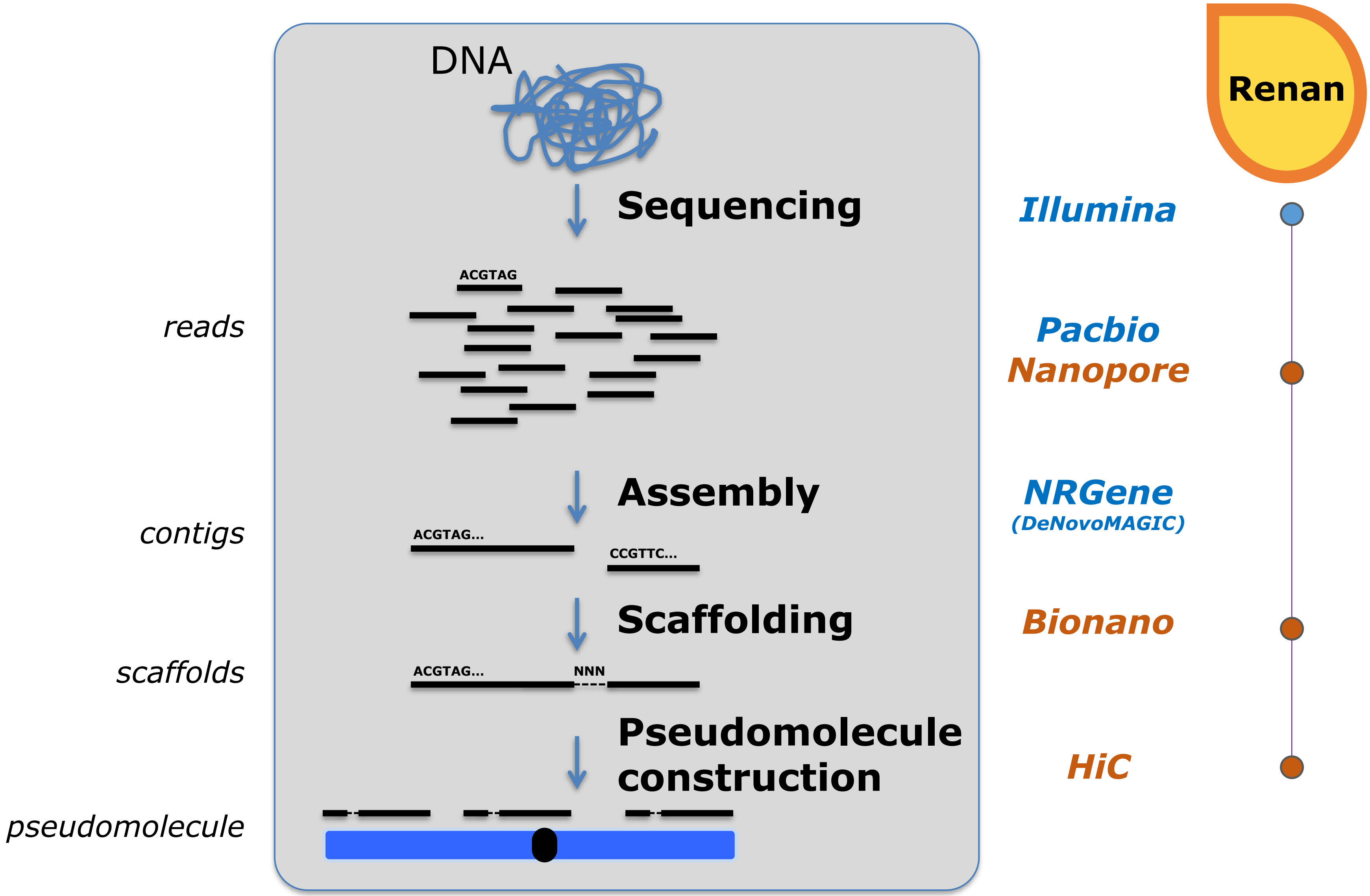


*Ae. ventricosa*



*G. Doussinault, J. Jahier, M. Bernard et al. (1980's 90's 00's)*

# ❑ Genome assembly: steps and techno



**Renan**

*reads*

**Sequencing**

***Illumina***

***Pacbio***
***Nanopore***

*contigs*

**Assembly**

***NRGene***
***(DeNovoMAGIC)***

*scaffolds*

**Scaffolding**

***Bionano***

*pseudomolecule*

**Pseudomolecule construction**

***HiC***

# ❑ *Triticeae* refseq current status (09/2022)

## Reference-quality genome assemblies

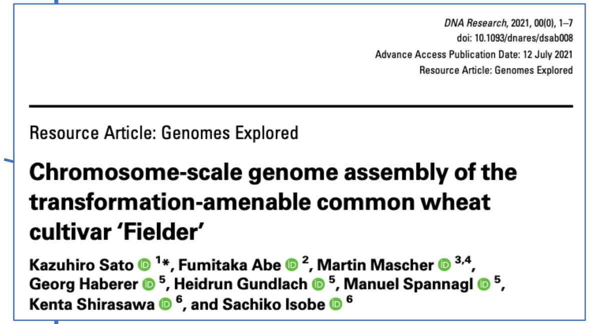
- *T. aestivum*
  - Chinese Spring (NRgene)
  - Arina*LrFor* Jagger Julius CDC_Lancer CDC_Landmark Mace Norin61 Stanley SY-Mattis Spelta (NRGene)
  - Zang1817 (NRGene)
  - Fielder (Pacbio HiFi)
  - Kariega (Pacbio HiFi)
  - Renan (ONT)
  - Attraktion (Pacbio HiFi)
- *T. urartu* AA
- *T. durum* AABB
- *T. dicoccoides* AABB
- *Ae. tauschii* DD
- *Ae. searsii* DD
- *Ae. bicornis* DD
- *Ae. longissima* DD
- *Ae. sharonensis* DD
- *Ae. speltoides* SS

Article | Open Access | Published: 08 October 2020

**Origin and adaptation to high altitude of Tibetan semi-wild wheat**

Weilong Guo, Mingming Xin, Zihao Wang, Yingyin Yao, Zhaorong Hu, Wanjun Song, Kuohai Yu, Yongming Chen, Xiaobo Wang, Panfeng Guan, Rudi Appels, Huiru Peng ✉, Zhongfu Ni ✉ & Qixin Sun ✉

*Nature Communications* **11**, Article number: 5085 (2020) | Cite this article

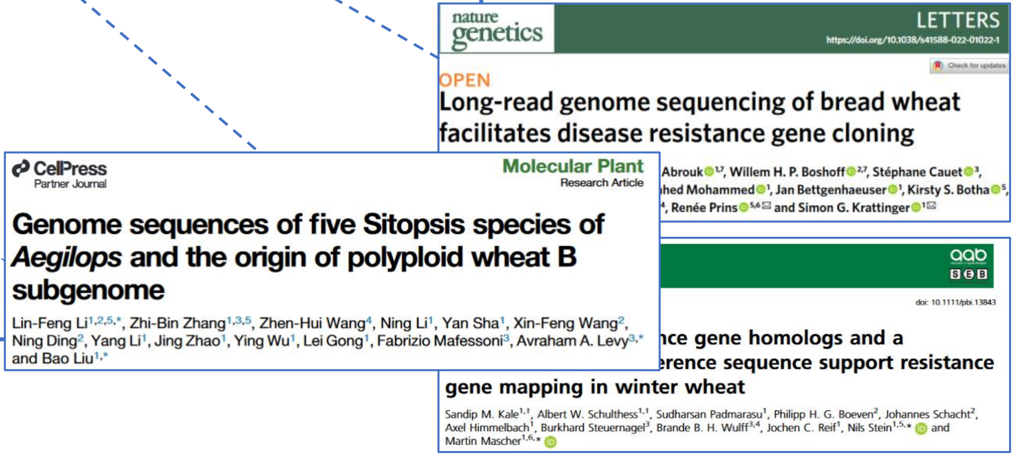


*DNA Research*, 2021, 00(0), 1–7
doi: 10.1093/dnares/dsab008
Advance Access Publication Date: 12 July 2021
Resource Article: Genomes Explored

Resource Article: Genomes Explored

**Chromosome-scale genome assembly of the transformation-amenable common wheat cultivar 'Fielder'**

**Kazuhiro Sato[1*], Fumitaka Abe[2], Martin Mascher[3,4], Georg Haberer[5], Heidrun Gundlach[5], Manuel Spannagl[5], Kenta Shirasawa[6], and Sachiko Isobe[6]**



nature genetics

LETTERS

https://doi.org/10.1038/s41588-022-01022-1

Check for updates

OPEN

**Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning**

Abrouk[1,7], Willem H. P. Boshoff[2,7], Stéphane Cauet[3], hed Mohammed[1], Jan Bettgenhaeuser[1], Kirsty S. Botha[5], [4], Renée Prins[5,6] ✉ and Simon G. Krattinger[1] ✉

CellPress
Partner Journal

Molecular Plant
Research Article

**Genome sequences of five Sitopsis species of *Aegilops* and the origin of polyploid wheat B subgenome**

Lin-Feng Li[1,2,5,*], Zhi-Bin Zhang[1,3,5], Zhen-Hui Wang[4], Ning Li[1], Yan Sha[1], Xin-Feng Wang[2], Ning Ding[2], Yang Li[1], Jing Zhao[1], Ying Wu[1], Lei Gong[1], Fabrizio Mafessoni[3], Avraham A. Levy[3,*] and Bao Liu[1,*]

aab
S E B

doi: 10.1111/pbi.13843

**nce gene homologs and a erence sequence support resistance gene mapping in winter wheat**

Sandip M. Kale[1,†], Albert W. Schulthess[1,†], Sudharsan Padmarasu[1], Philipp H. G. Boeven[2], Johannes Schacht[2], Axel Himmelbach[1], Burkhard Steuernagel[3], Brande B. H. Wulff[3,4], Jochen C. Reif[1], Nils Stein[1,5,*] and Martin Mascher[1,6,*]

GigaScience, 2022, 11, 1–18
DOI: 10.1093/gigascience/giac034
RESEARCH

# Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding

Jean-Marc Aury[1,*], Stefan Engelen[1], Benjamin Istace[1], Cécile Monat[2], Pauline Lasserre-Zuber[2], Caroline Belser[1], Corinne Cruaud[3], Hélène Rimbert[2], Philippe Leroy[2], Sandrine Arribat[4], Isabelle Dufau[4], Arnaud Bellec[4], David Grimbichler[5], Nathan Papon[2], Etienne Paux[2], Marion Ranoux[2], Adriana Alberti[1,6], Patrick Wincker[1] and Frédéric Choulet[2,*]

[1] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France
[2] GDEC, Université Clermont Auvergne, INRAE, UMR1095, 63000 Clermont-Ferrand, France
[3] Commissariat à l'Energie Atomique (CEA), Institut François Jacob, Genoscope, F-91057 Evry, France
[4] INRAE, CNRGV French Plant Genomic Resource Center, F-31320, Castanet Tolosan, France
[5] Mésocentre Clermont Auvergne, DOSI / Bâtiment Turing, 7 avenue Blaise Pascal, 63178 Aubière, France
[6] Present address: Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France
[*] Correspondence address. Jean-Marc Aury, Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France. E-mail: jmaury@genoscope.cns.fr; Frédéric Choulet, GDEC, Université Clermont Auvergne, INRAE, UMR1095, 63000 Clermont-Ferrand, France. E-mail: frederic.choulet@inrae.fr

## Abstract

**Background:** The sequencing of the wheat (*Triticum aestivum*) genome has been a methodological challenge for many years owing to its large size (15.5 Gb), repeat content, and hexaploidy. Many initiatives aiming at obtaining a reference genome of cultivar Chinese Spring have been launched in the past years and it was achieved in 2018 as the result of a huge effort to combine short-read sequencing with many other resources. Reference-quality genome assemblies were then produced for other accessions, but the rapid evolution of sequencing technologies offers opportunities to reach high-quality standards at lower cost.

**Results:** Here, we report on an optimized procedure based on long reads produced on the Oxford Nanopore Technology PromethION device to assemble the genome of the French bread wheat cultivar Renan.

**Conclusions:** We provide the most contiguous chromosome-scale assembly of a bread wheat genome to date. Coupled with an annotation based on RNA-sequencing data, this resource will be valuable for the crop community and will facilitate the rapid selection of agronomically important traits. We also provide a framework to generate high-quality assemblies of complex genomes using ONT.

**Keywords:** wheat, hexaploid genome, long-reads, nanopore sequencing, genome assembly, haplotype characterization, introgressions

# ❑ *Step#1: sequencing*

NovaSeq
**30x**

MinION 2 runs
PromethION 18 runs

- Total 63x
- Select **30x** with longer reads
  (reads>50kb ➔ 15x)

**Renan**

***Illumina***

***Nanopore***

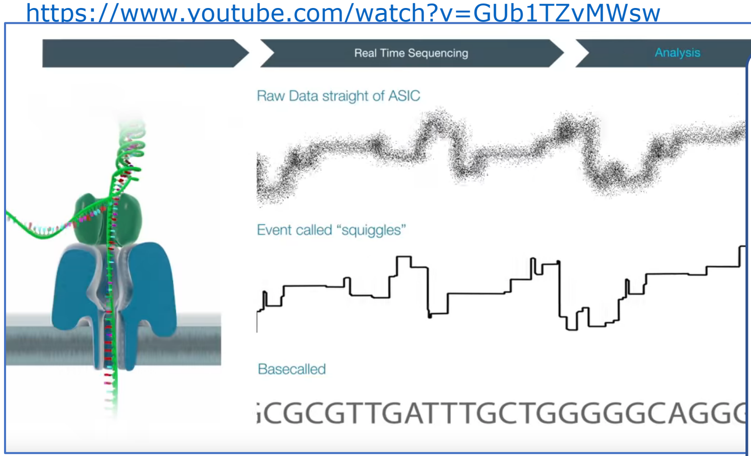
***Bionano***

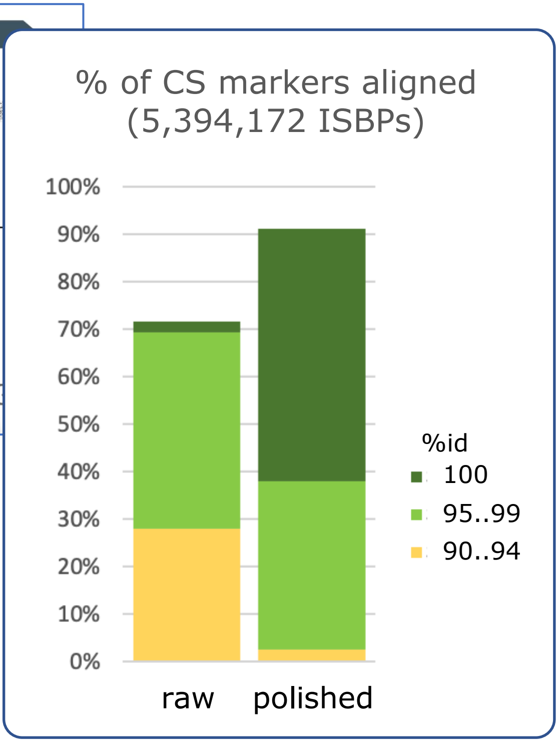
***HiC***

# ❑ Step#2: read assembly

**Renan**

https://www.youtube.com/watch?v=GUb1TZvMWsw




% of CS markers aligned (5,394,172 ISBPs)

***Nanopore***

Early stages
-> error rate in reads: 12%
-> error rate in assembly: 4%

- basecalling
- polishing

Illumina
30x

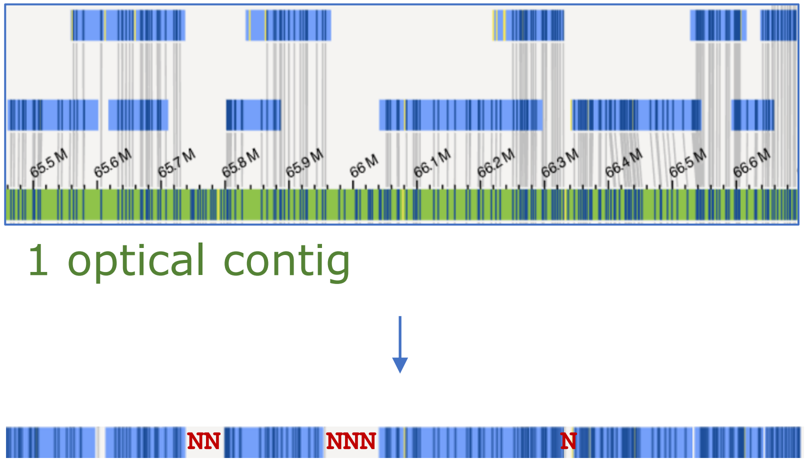
ACGTAGACGTACGGCTGCATCGCATTCATCGGCGGGGGGGGGCTGCTCGTCGCTCCTCGT

assembly

# ☐ *Step#3: scaffolding*

**Renan**

11 sequence contigs



1 optical contig

1 scaffold

*Illumina*

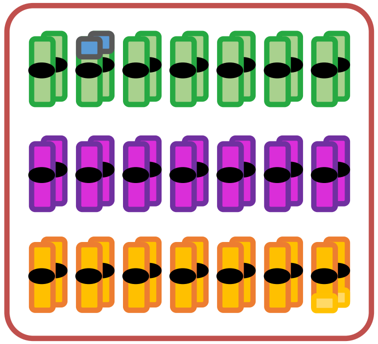
*Nanopore*

***Bionano***

*HiC*

# ❑ *Assembly metrics*

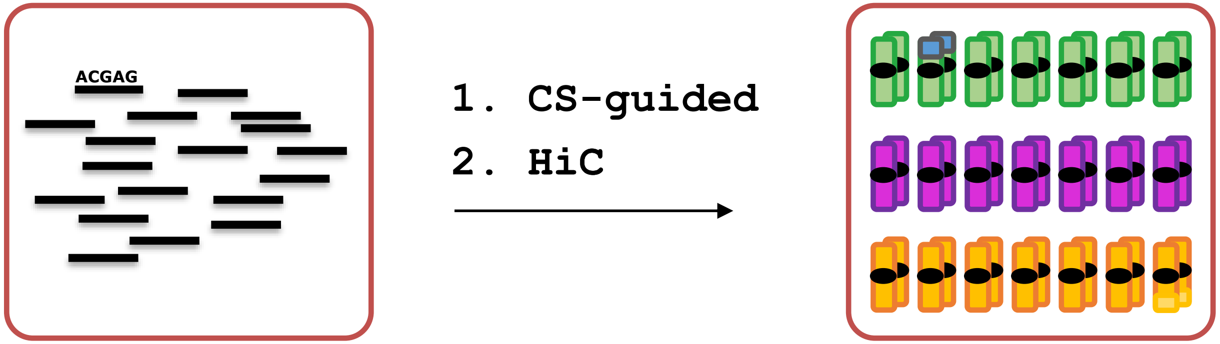
- ❖ **Metrics**
  - ➢ Genome size=14.3 Gb
  - ➢ 2904 scaffolds (=138/chr)



- ❖ **Quality Assessment**

  Genes: 98% CS genes found
  TEs: 92% CS ISBPs found
  QV: lower (99.95% accuracy) than NRGene-based assemblies (99.99%)

# ❑ ***Step#4: pseudomolecule construction***

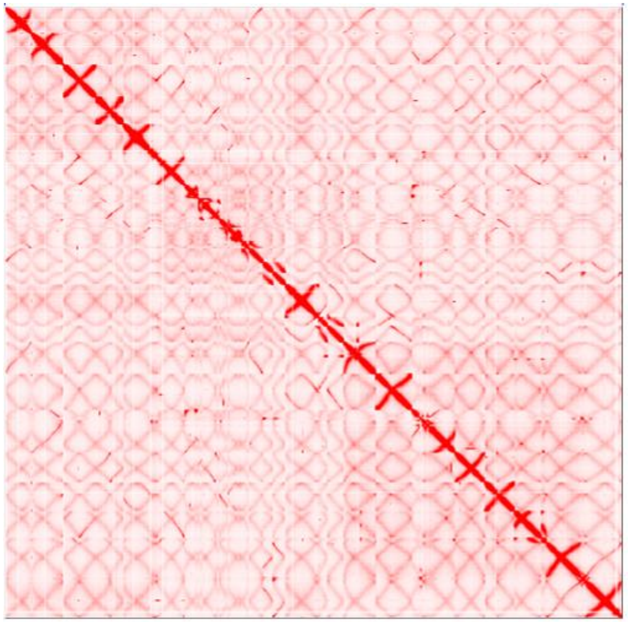


```
1. CS-guided
2. HiC
```

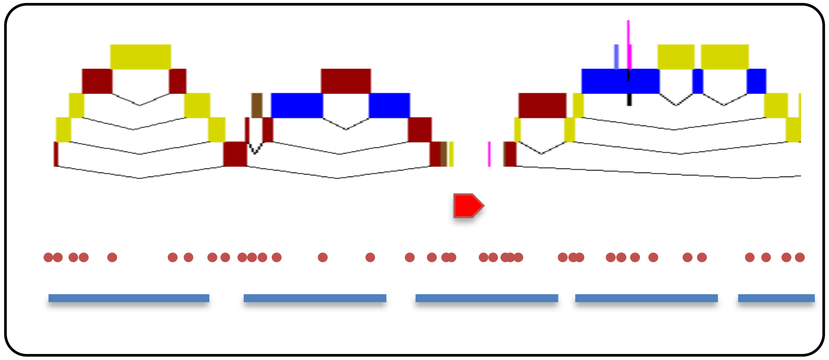
2904 scaff

21 chromosomes

X not efficient enough to build a chromosome-scale assembly



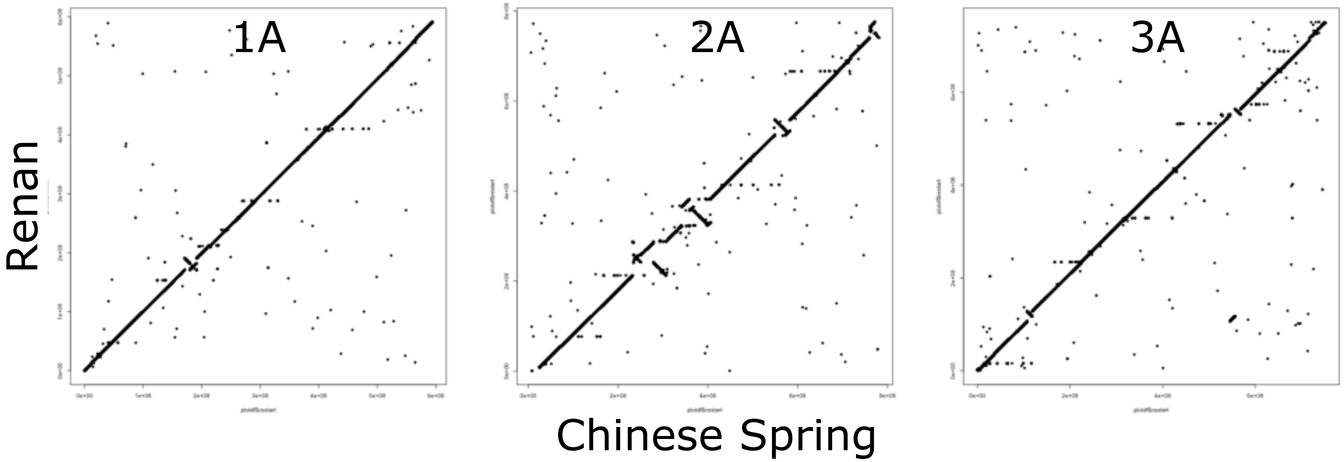
# ❑ *Step#4-1: CS-guided pseudomol construction*


TEs

Genes (1/130kb)

ISBPs (1/3kb)

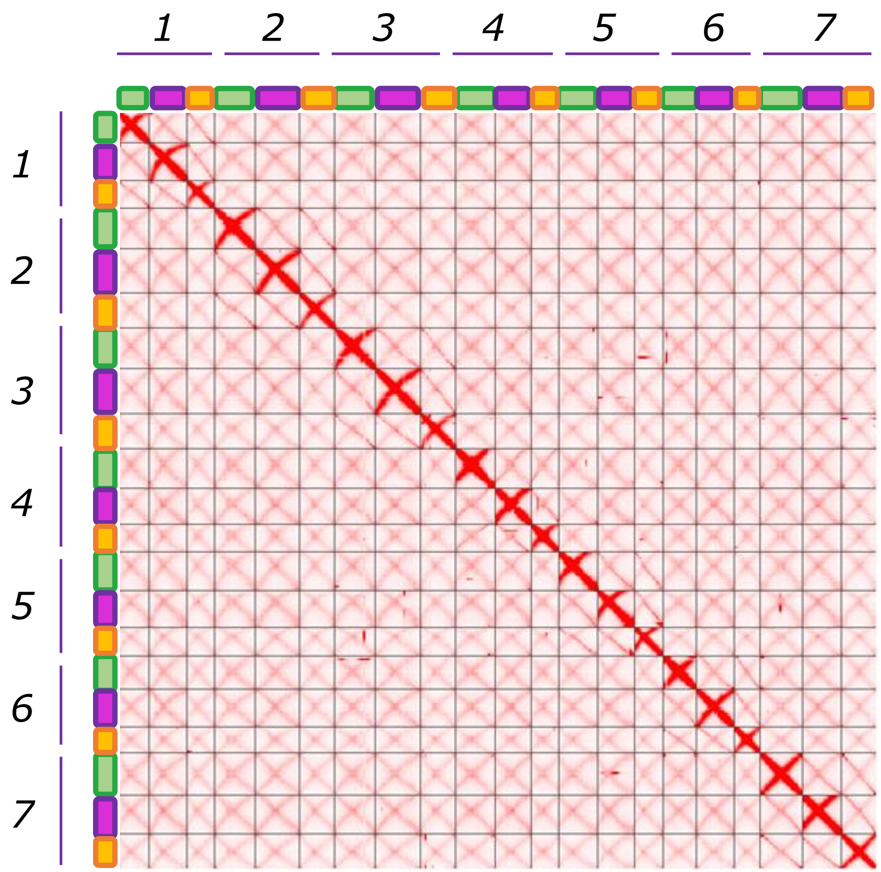
scaffolds



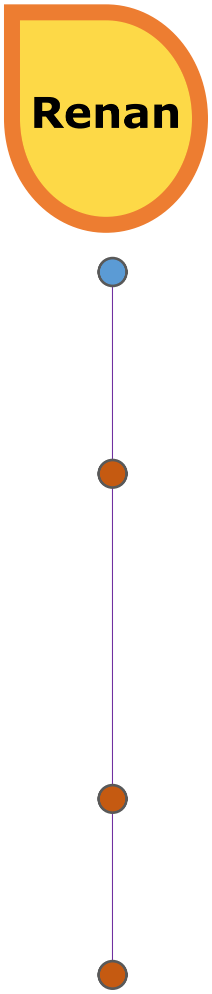
⇒ 14.2 Gb (0.1 Gb unanchored)

❑ ***Step#4-1: CS-guided pseudomol construction***

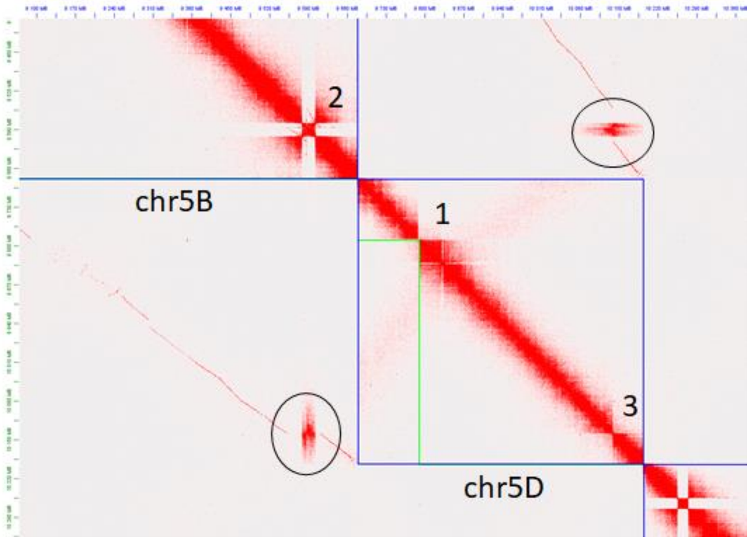
❑ ***Step#4-2: HiC corrections***

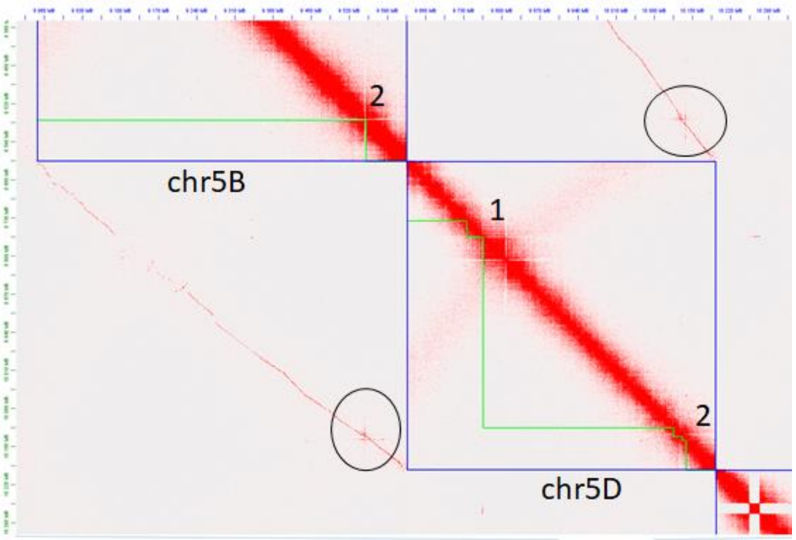


***HiC***



## *Step#4-2: HiC corrections*

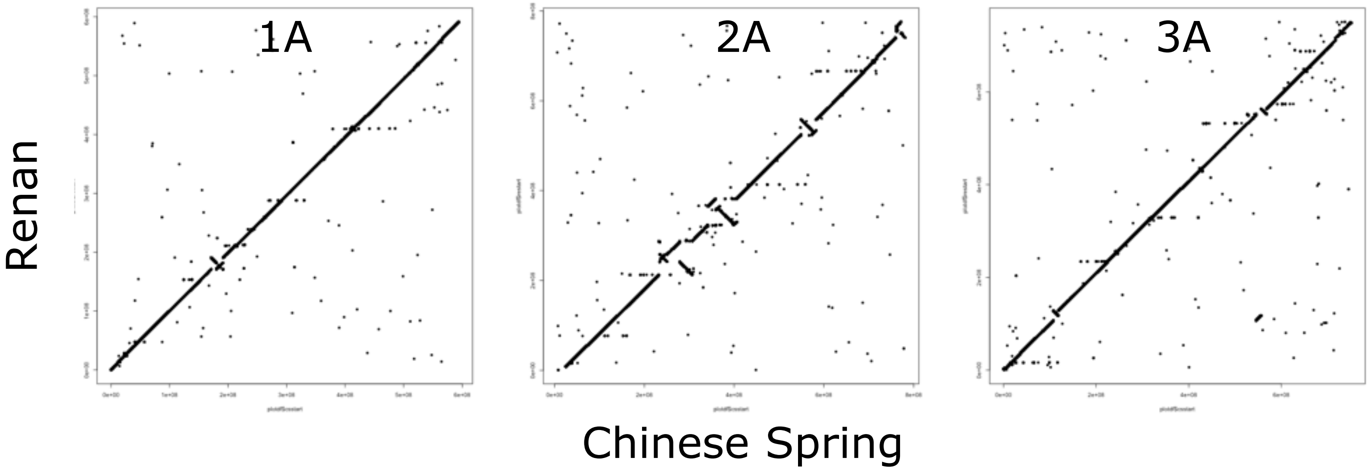




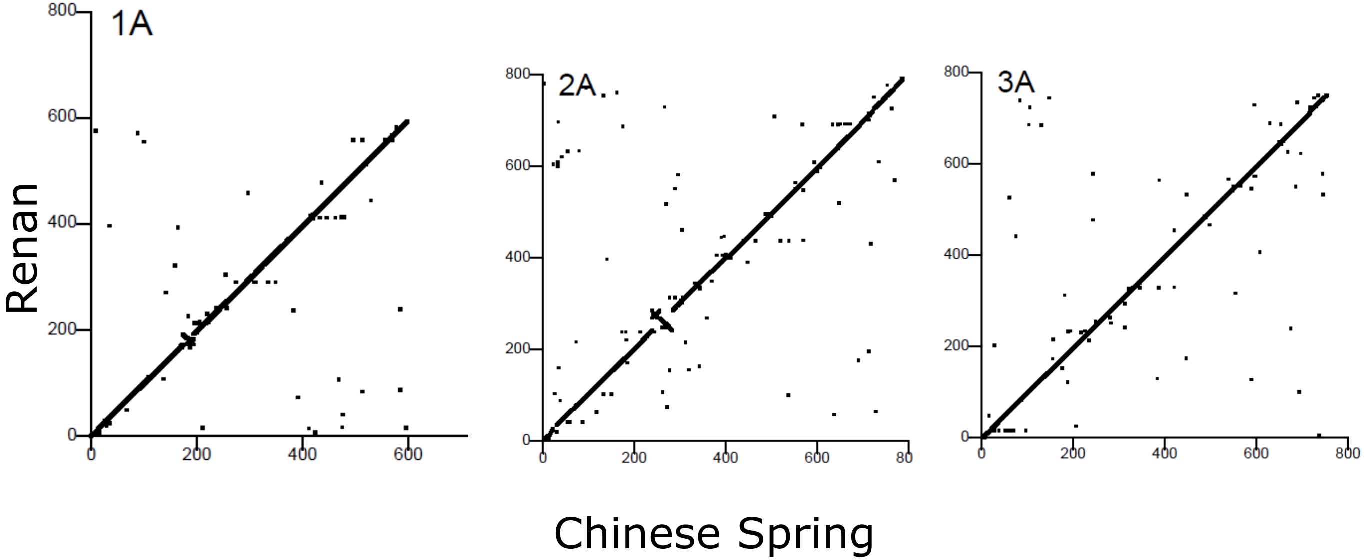
- **18** chimeric scaff
- **198** scaff with correction of location and/or orientation

↓

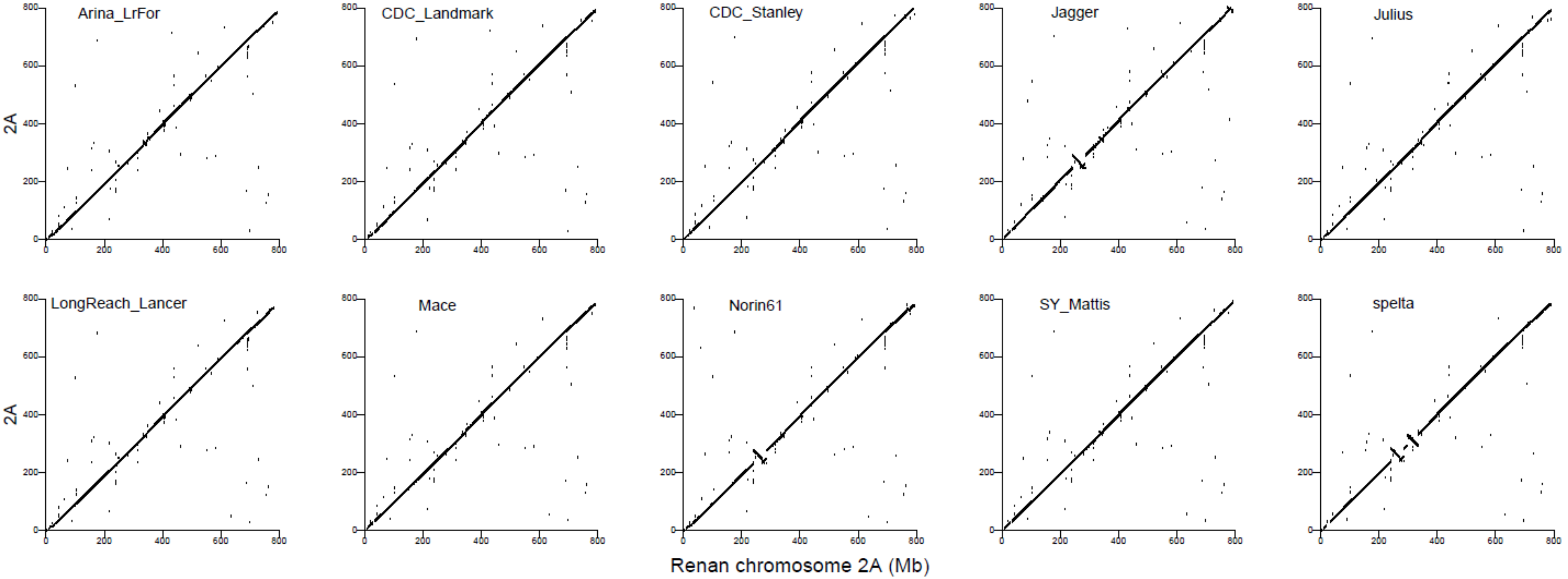
21 chr + chrUn 61 Mb (<1%)

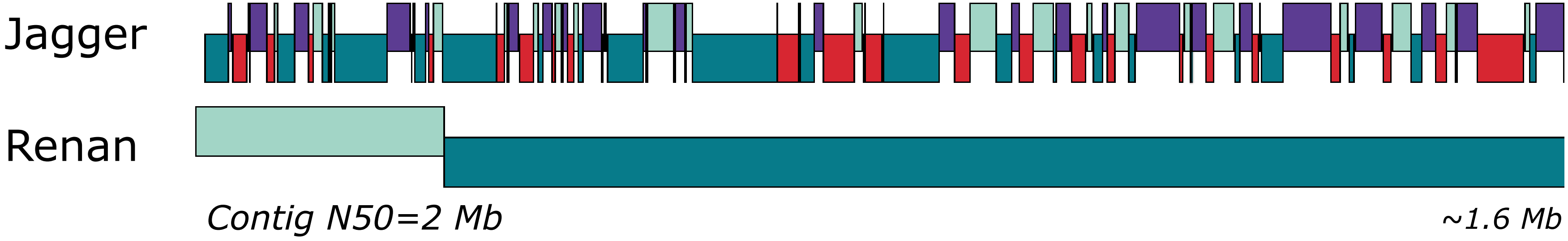


After HiC-based corrections

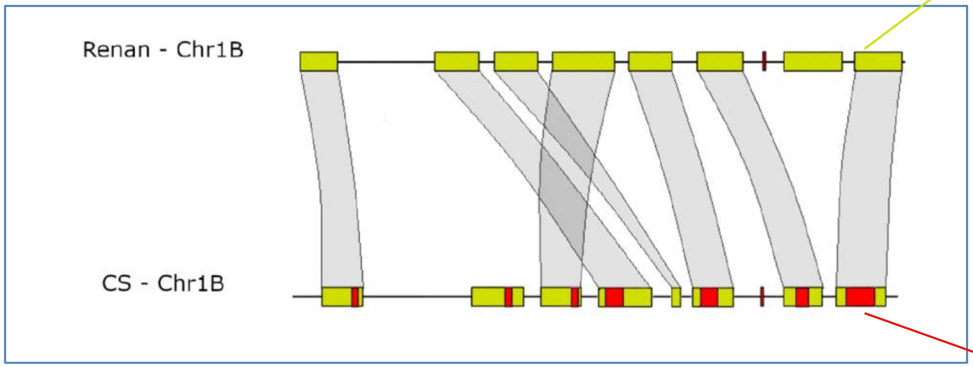


# Renan Refseq_v2 *versus* 10 wheat genomes





ω-gliadin gene cluster

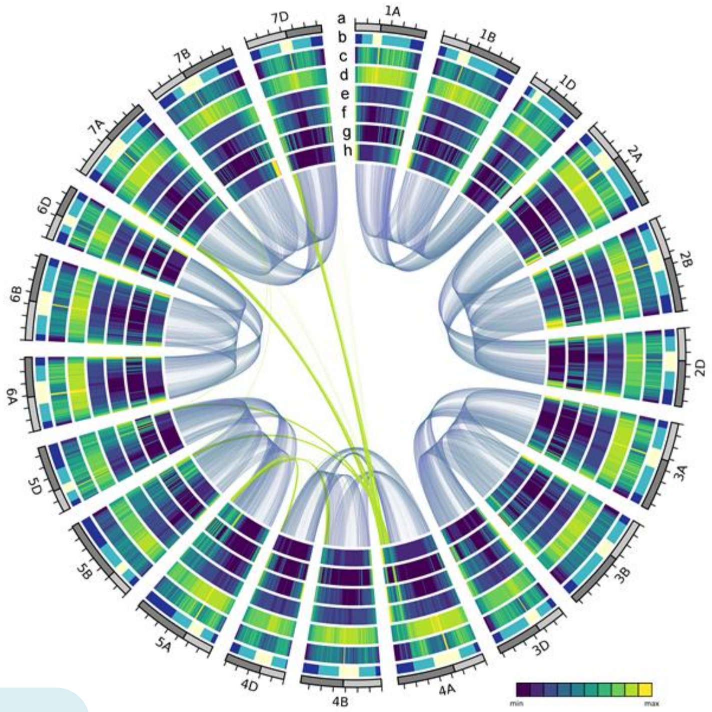


ATGAAGACCTTCATCATATTTGTCCTCCTTGCTATGGTGATGAGCATCGCCAGTGCCACTAGGCAACTAA
GCCCTAGAAGCAAGGAGTTGCAAACACCACAAGAACAATTCCCCCAACAGCAATTTCCTCAACCACAACA
AATCCTC**CAACAACAATTCCCCCAACAACAACAAATCCCCCAGCAACCGCAACAATTCCCCCAACAACAA**
**ATCCCGCAACAGCAAATCCCGCAACAACAAATCCCCCAACAACAACAAATCCCCCAGCAACCACAACAAT**
**TCCCCCAACAACAGCAATTCCCCCAACAACAACAAATCCCCCAGCAACCACAACAATTCCCCCAACAACA**
**GCAATTCCCCCAACAACAACAAATCCTCCAGCAACCACAACAATTCCCCCAACAACAGCAATTCCCCCAA**
**ACAATTACCCCAACAACAATACCCCGAACA // AACAATTCCCCCAACAGCAATTCCCCCAACAACAAC**
**AATTTCCCCAACAACACCAGTTACCGCAACAACAATTCCCCCAACAACAACAAATCTCCGAGCAACCACA**
**ACAATTCCCCCAACAACAACAATTCCCCCAACAACAACAATTCCCCCAACAGCAATTCCCGCAACAACAA**
**TTTCCCCAACAACAACAATTCCCACAACAACAACAGTTCCCTCAACAACAACAATTCCCGCAACAACAAA**
**TCCCTGAACAACAA**TAATTCCCCCAACAACAATTCCCCCAGCAACCATCACAACAATTTCCCCAACAACA
ATTTCCAATACCATACCCACCCCAACAATCACAAGAACCTTCCCCATACCAACAATATCCACAACAACAA
CCATATGGGAGCAACGTTATAAGTATCAGTGGCCGATGA

ATGAAGACCTTCATCATATTTGTCCTCCTTGCTATGGTGATGAGCATCGCCAGTGCCACTAGGCAACTAA
GCCCTAGAAGCAAGGAGTTGCAAACACCACAAGAACAATTCCCCCAACAGCAATTTCCTCAACCACAACA
AATCCTCCAACAACAATTCCCCCAACAACAACAAATCCCCCAGCAACCGCAACAATTCCCCCAACAACAA
ATCCCGCAACAGCAAATCCCGCAACAACAAATCCCCCAACAACAACNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNN // NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNACAACAATTCCCCCAACAACCACAACAACAATTCCCCCAACAACAATTTCCCATACCATACC
CACCCCAACAACAATCACAAGAAGCTTCCCCATACCAACAATACCACAACAACAACTATCTGGGAGCGAC
GTTATAA

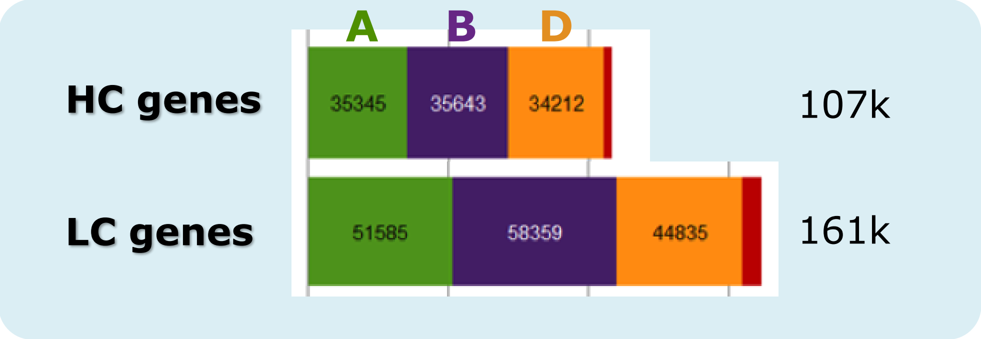
# ○ Predicting genes is still NOT routine

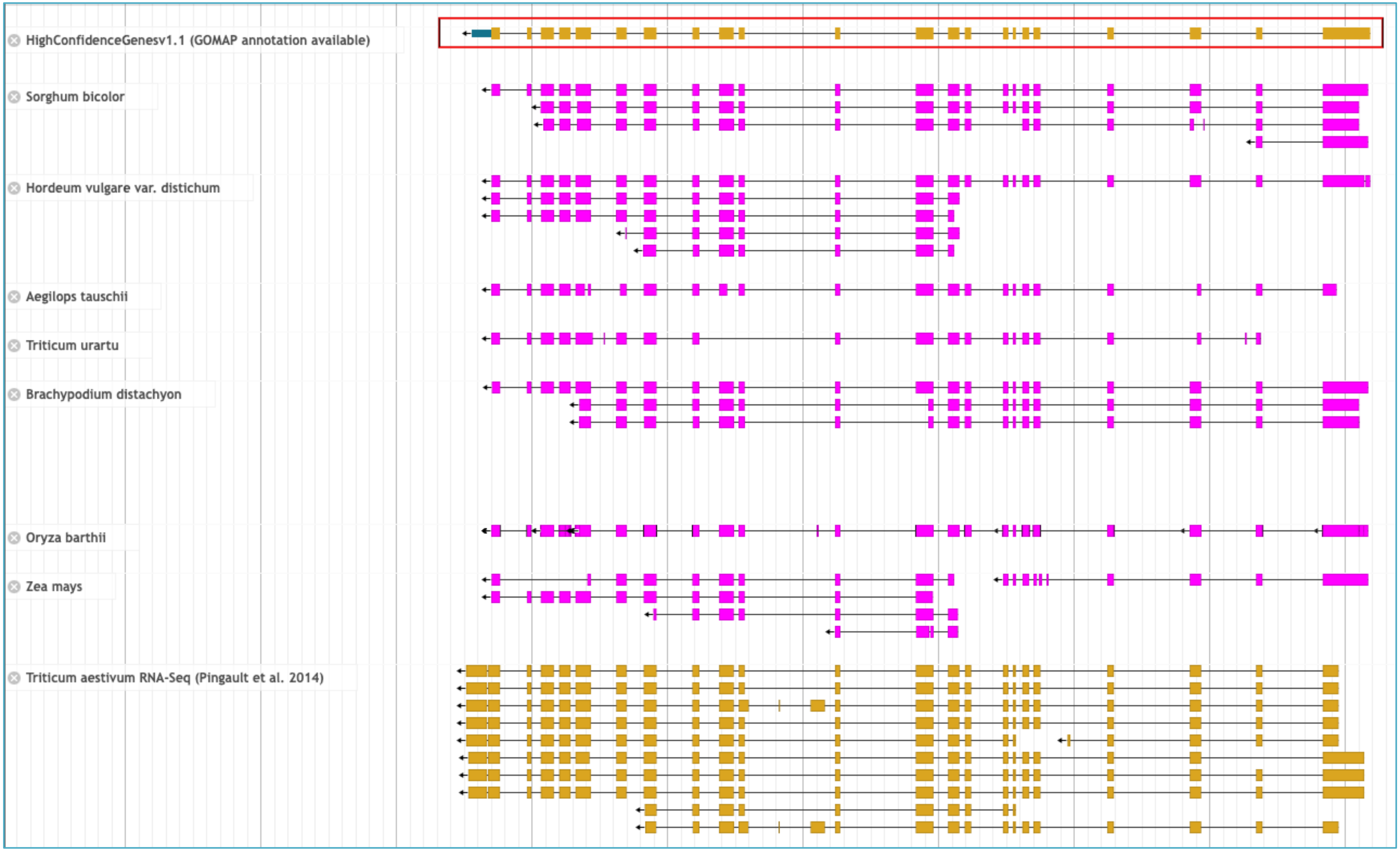
**GDEC** Clermont, France – *Rimbert Leroy Choulet et al.*
**PGSB** Munich, Germany – *Spannagl Twardziok et al.*
**EI** Norwich, UK – *Swarbreck Venturini et al.*

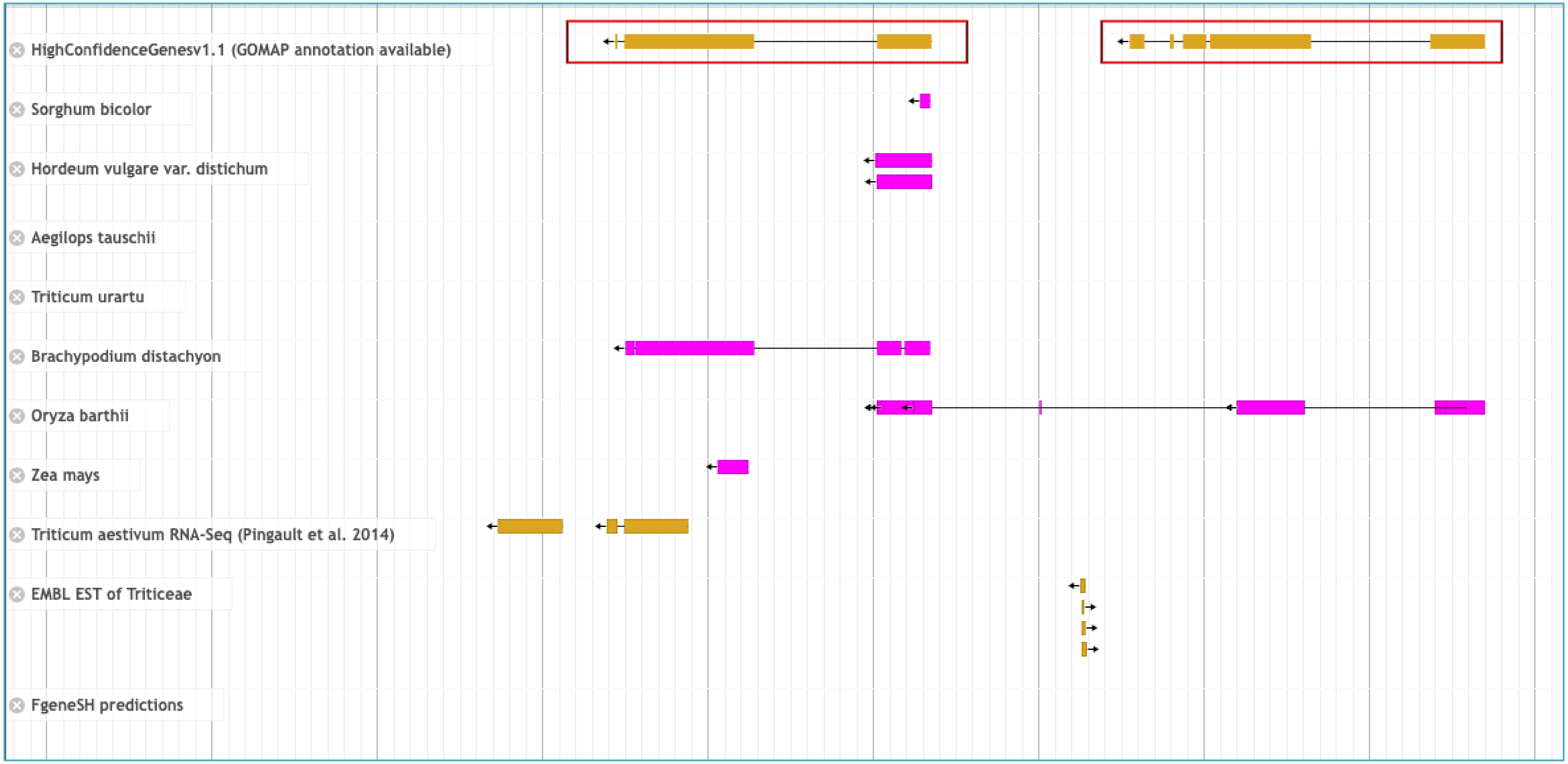


## IWGSC RefSeq (Chinese Spring)

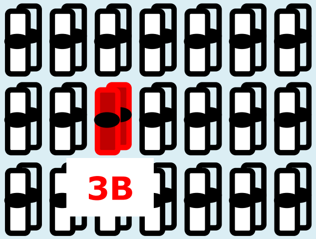
- 107,891 genes




HighConfidenceGenesv1.1 (GOMAP annotation available)
Sorghum bicolor
Hordeum vulgare var. distichum
Aegilops tauschii
Triticum urartu
Brachypodium distachyon
Oryza barthii
Zea mays
Triticum aestivum RNA-Seq (Pingault et al. 2014)

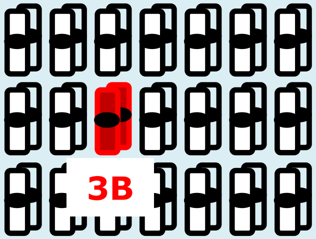

HighConfidenceGenesv1.1 (GOMAP annotation available)
Sorghum bicolor
Hordeum vulgare var. distichum
Aegilops tauschii
Triticum urartu
Brachypodium distachyon
Oryza barthii
Zea mays
Triticum aestivum RNA-Seq (Pingault et al. 2014)
EMBL EST of Triticeae
FgeneSH predictions

chr3B BAC-by-BAC



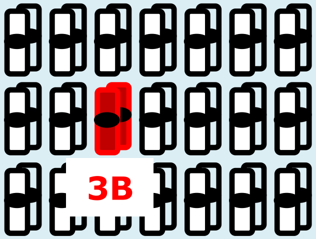
*Choulet et al. 2014*

TGAC_v1



*Clavijo et al. 2017*

Chr. Survey Seq



*IWGSC 2014*

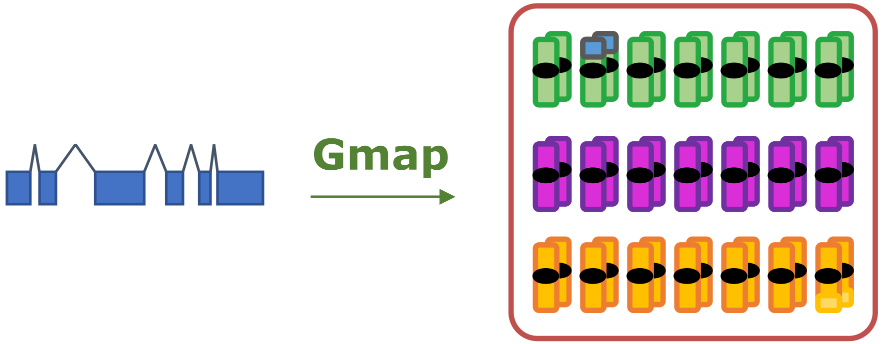
7264 genes

5728 genes
(**26**% identical)

5862 genes
(**12**% identical)

➢ *denovo* gene annotation not adapted to pangenomics

# ○ Gene projections / Annotation transfer



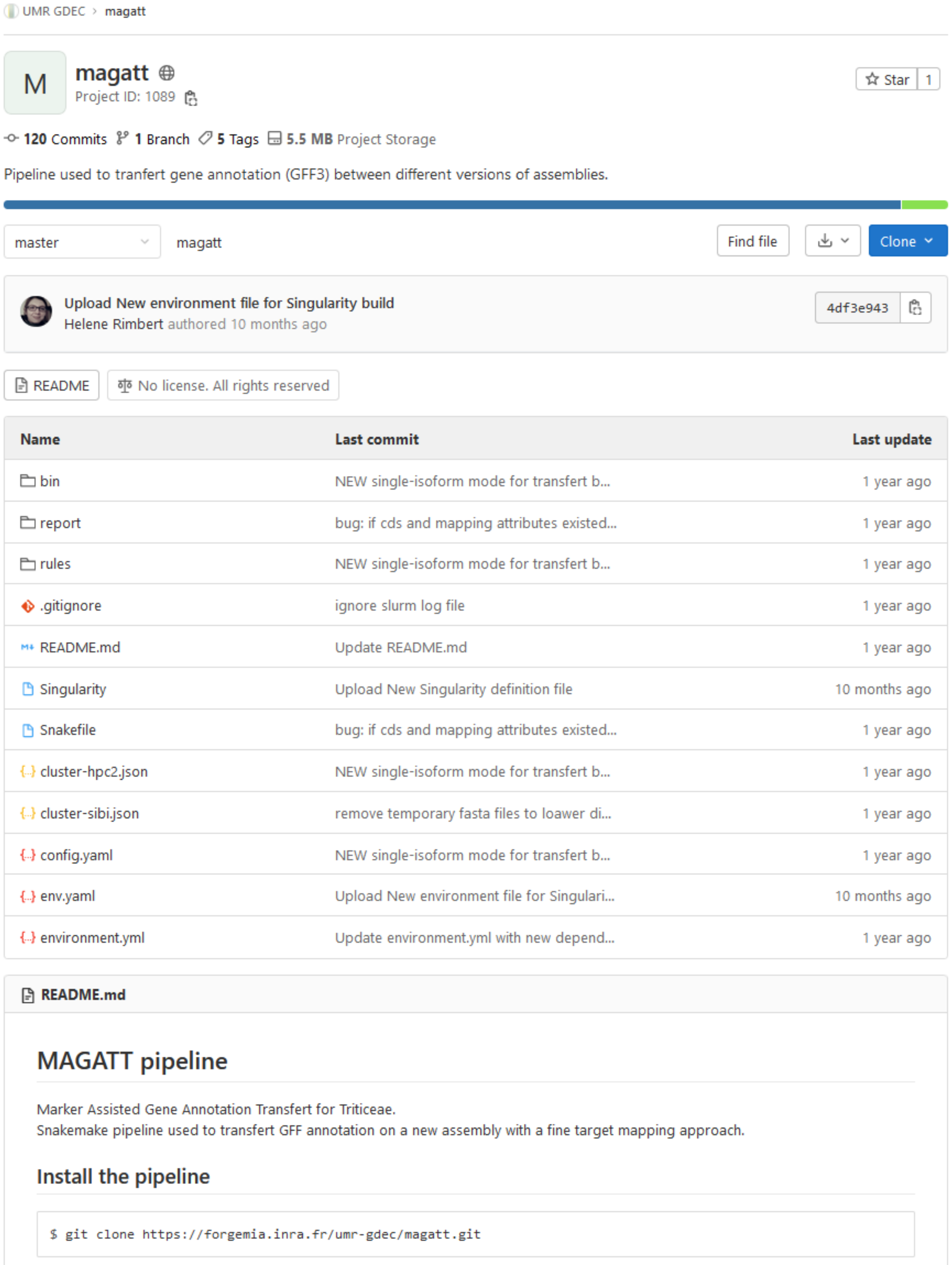
Chinese Spring

Reference annotation

Renan

**~12% ambiguous mapping**

- *denovo* gene annotation not adapted to pangenomics
- Gmap *vs* whole genome introduces mis-alignment errors
  - Develop **MAGATT** pipeline



UMR GDEC › magatt

M

**magatt**

Project ID: 1089

☆ Star 1

120 Commits · 1 Branch · 5 Tags · 5.5 MB Project Storage

Pipeline used to tranfert gene annotation (GFF3) between different versions of assemblies.

master · magatt · Find file · Clone

**Upload New environment file for Singularity build**
Helene Rimbert authored 10 months ago

4df3e943

README · No license. All rights reserved

| Name | Last commit | Last update |
|---|---|---|
| bin | NEW single-isoform mode for transfert b... | 1 year ago |
| report | bug: if cds and mapping attributes existed... | 1 year ago |
| rules | NEW single-isoform mode for transfert b... | 1 year ago |
| .gitignore | ignore slurm log file | 1 year ago |
| README.md | Update README.md | 1 year ago |
| Singularity | Upload New Singularity definition file | 10 months ago |
| Snakefile | bug: if cds and mapping attributes existed... | 1 year ago |
| cluster-hpc2.json | NEW single-isoform mode for transfert b... | 1 year ago |
| cluster-sibi.json | remove temporary fasta files to loawer di... | 1 year ago |
| config.yaml | NEW single-isoform mode for transfert b... | 1 year ago |
| env.yaml | Upload New environment file for Singulari... | 10 months ago |
| environment.yml | Update environment.yml with new depend... | 1 year ago |

**README.md**

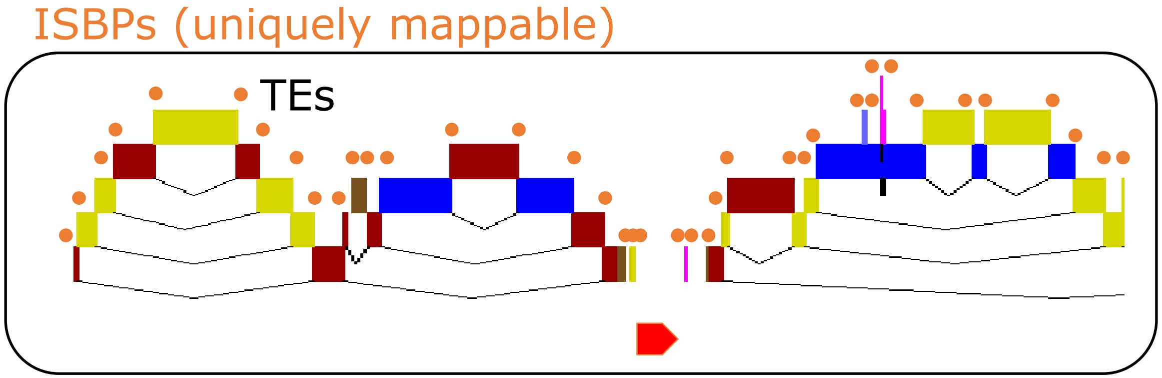
# MAGATT pipeline

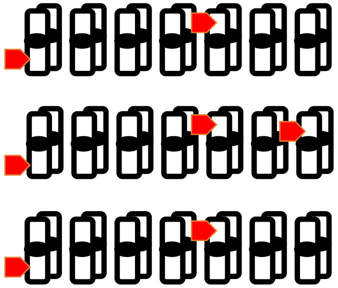
Marker Assisted Gene Annotation Transfert for Triticeae.
Snakemake pipeline used to transfert GFF annotation on a new assembly with a fine target mapping approach.

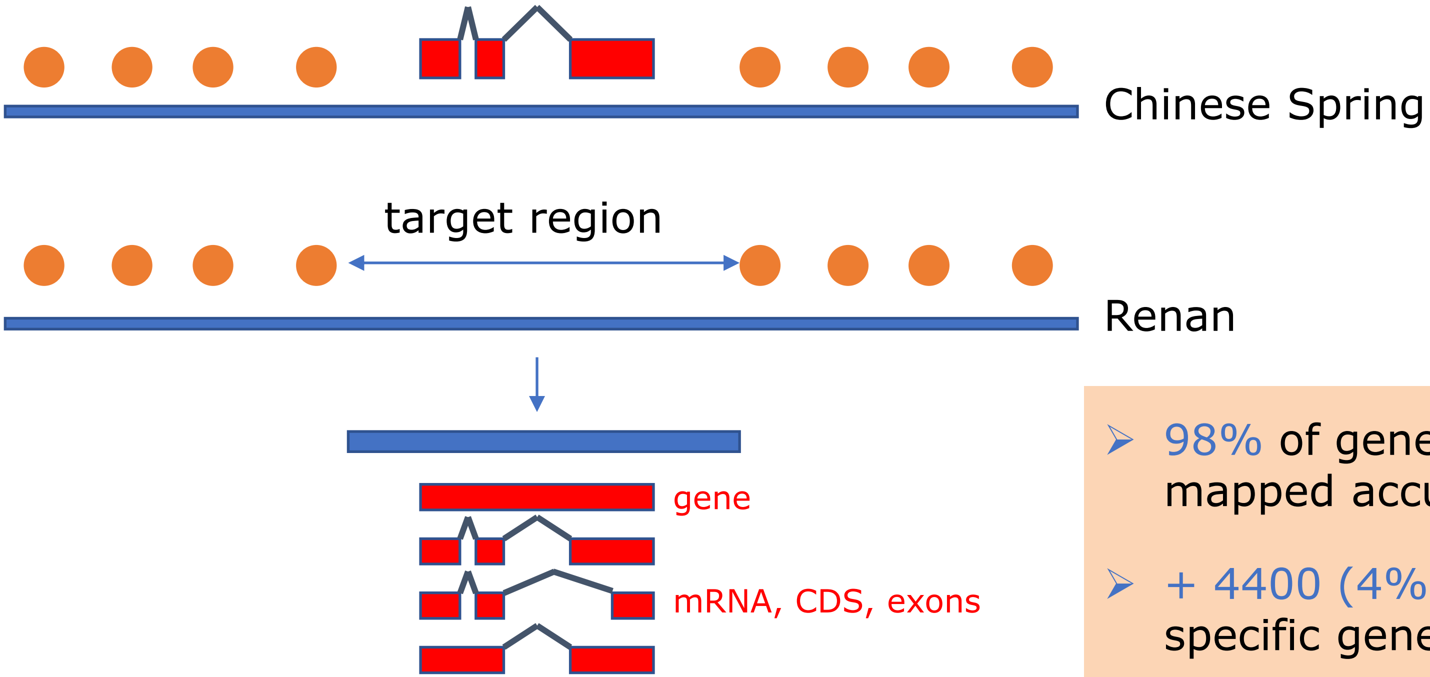
## Install the pipeline

```
$ git clone https://forgemia.inra.fr/umr-gdec/magatt.git
```

# ❑ **MAGATT** (**M**arker-**A**ssisted **G**ene **A**nnotation **T**ransfer for ***Triticeae***)

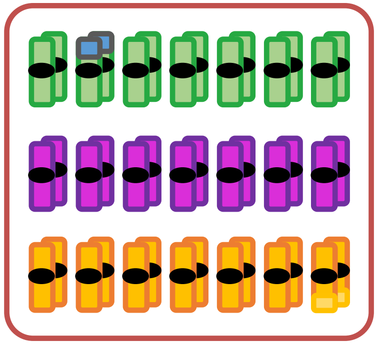


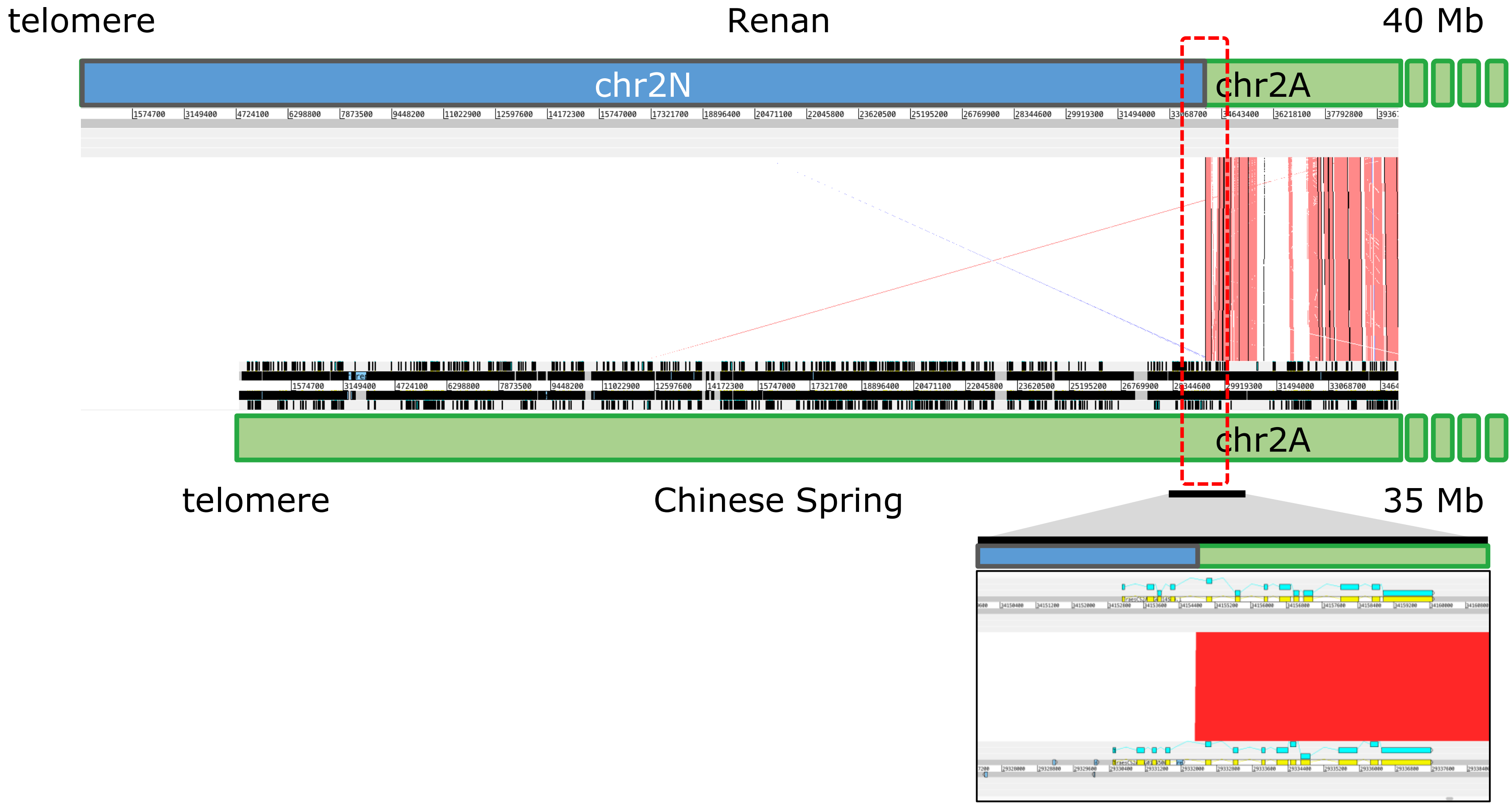


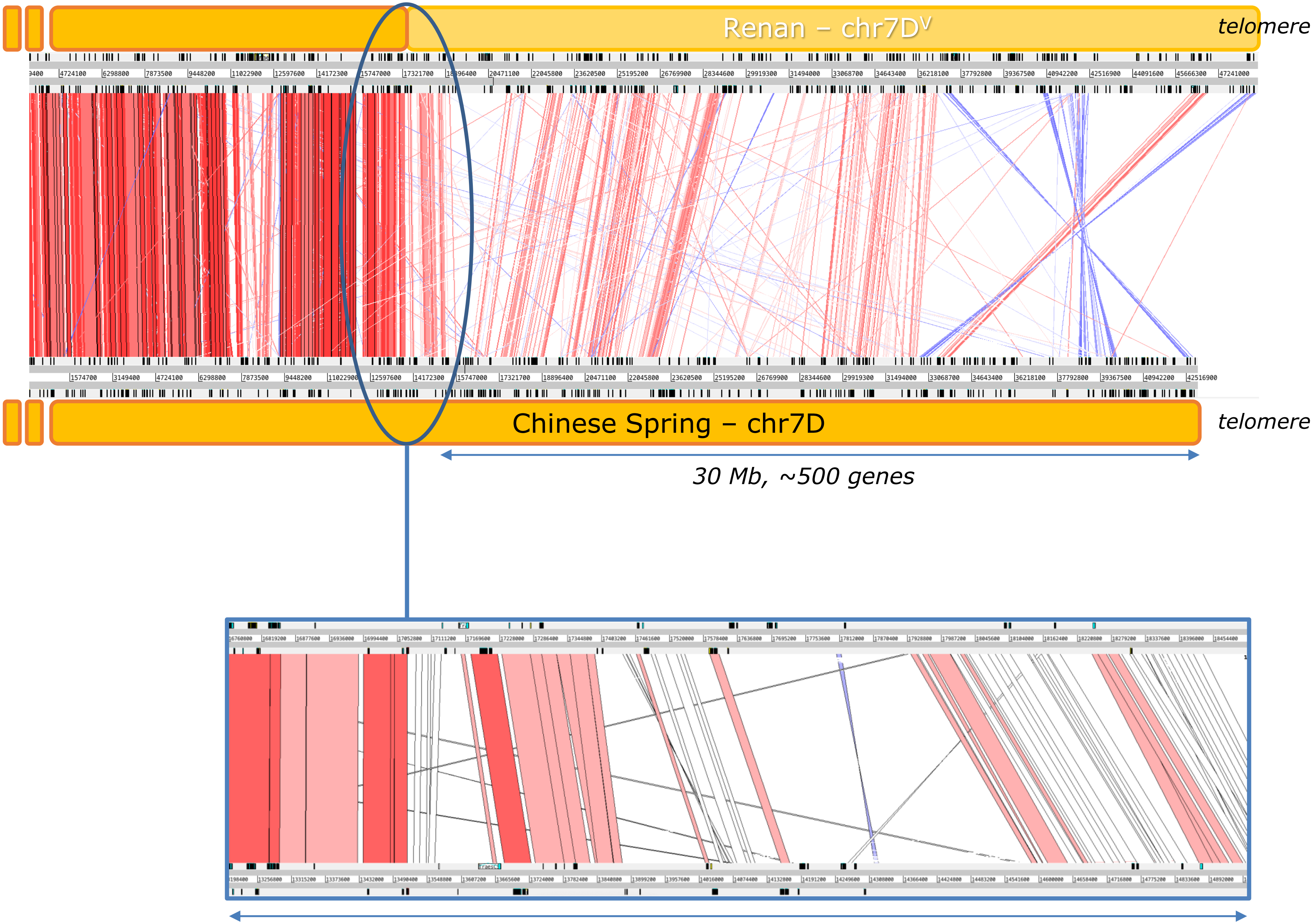


- ➢ 98% of genes mapped accurately
- ➢ + 4400 (4%) Renan specific genes

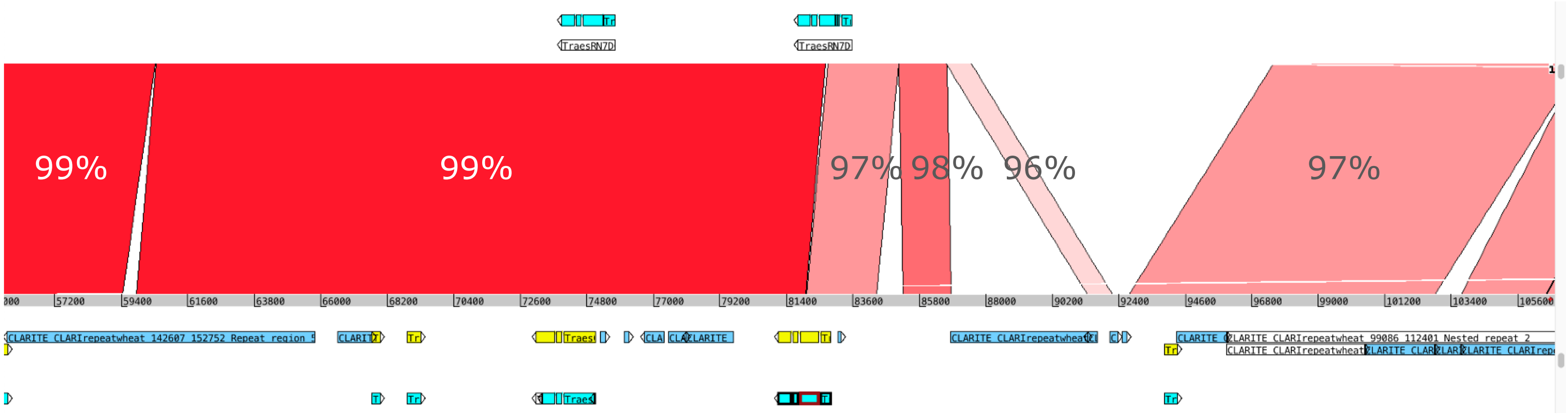
# ❑ *Introgressions*






Renan – chr7D$^V$
telomere
Chinese Spring – chr7D
telomere
30 Mb, ~500 genes

**Renan – chr7D** **Renan – chr7D$^V$**

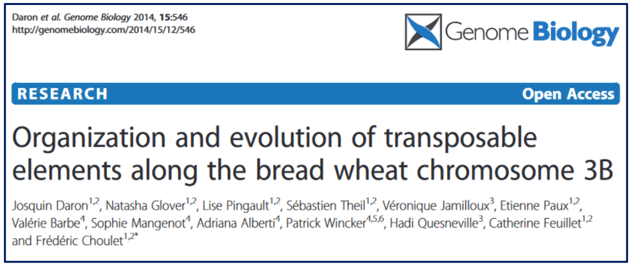


**Chinese Spring – chr7D**

# ❑ TE dynamics

- *PhD Nathan Papon (2019-2022)*

*Daron et al. 2015*



Daron *et al. Genome Biology* 2014, **15**:546
http://genomebiology.com/2014/15/12/546

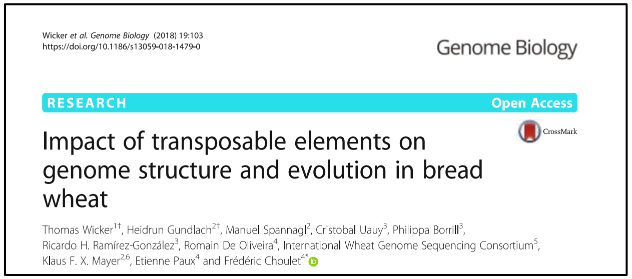
Genome **Biology**

**RESEARCH** **Open Access**

Organization and evolution of transposable elements along the bread wheat chromosome 3B

Josquin Daron[1,2], Natasha Glover[1,2], Lise Pingault[1,2], Sébastien Theil[1,2], Véronique Jamilloux[3], Etienne Paux[1,2], Valérie Barbe[4], Sophie Mangenot[4], Adriana Alberti[4], Patrick Wincker[4,5,6], Hadi Quesneville[3], Catherine Feuillet[1,2] and Frédéric Choulet[1,2*]

*Wicker et al. 2018*



Wicker *et al. Genome Biology* (2018) 19:103
https://doi.org/10.1186/s13059-018-1479-0

Genome Biology

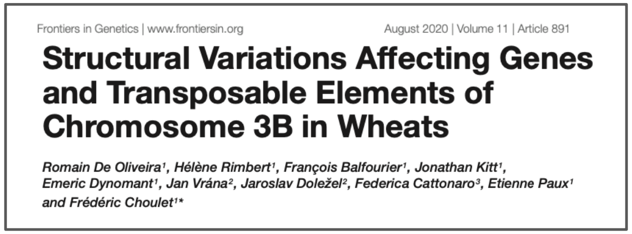
**RESEARCH** **Open Access**

CrossMark

Impact of transposable elements on genome structure and evolution in bread wheat

Thomas Wicker[1†], Heidrun Gundlach[2†], Manuel Spannagl[2], Cristobal Uauy[3], Philippa Borrill[3], Ricardo H. Ramírez-González[3], Romain De Oliveira[4], International Wheat Genome Sequencing Consortium[5], Klaus F. X. Mayer[2,6], Etienne Paux[4] and Frédéric Choulet[4*]

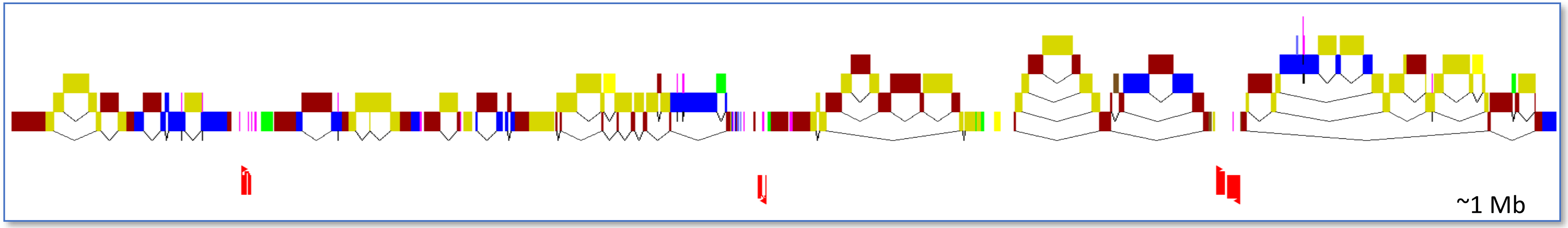
*De Oliveira et al. 2020*



Frontiers in Genetics | www.frontiersin.org August 2020 | Volume 11 | Article 891

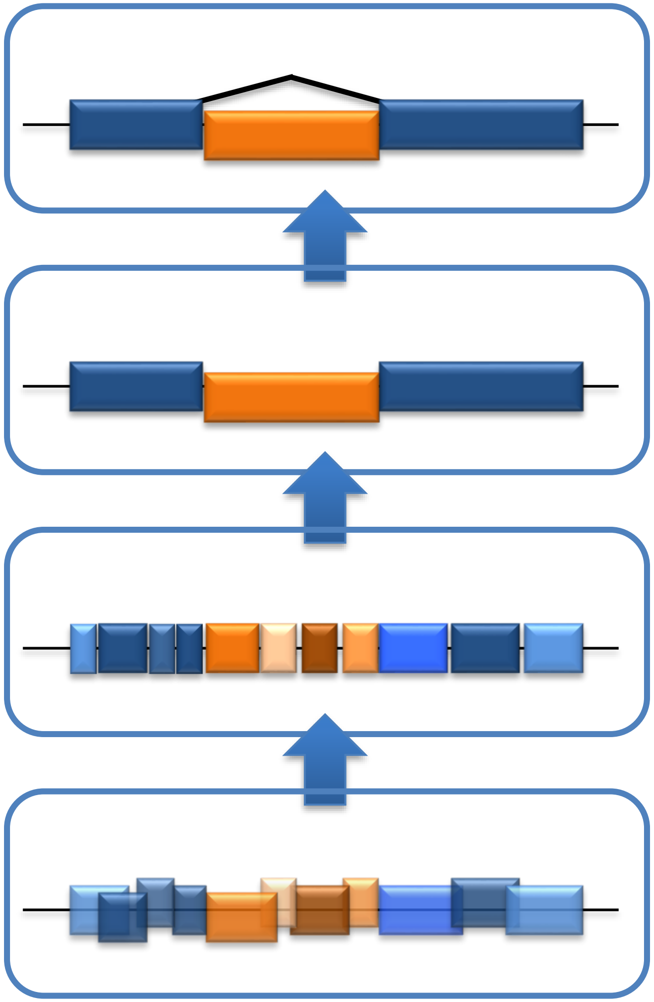
**Structural Variations Affecting Genes and Transposable Elements of Chromosome 3B in Wheats**

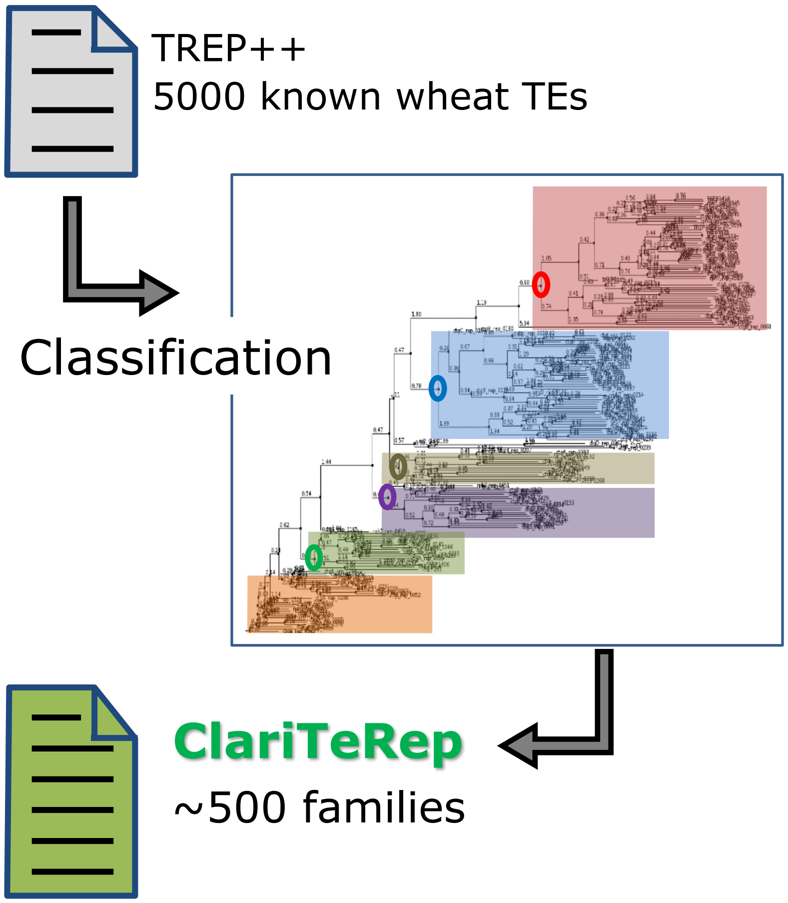
***Romain De Oliveira[1], Hélène Rimbert[1], François Balfourier[1], Jonathan Kitt[1], Emeric Dynomant[1], Jan Vrána[2], Jaroslav Doležel[2], Federica Cattonaro[3], Etienne Paux[1] and Frédéric Choulet[1*]***


~1 Mb

# ○ TE modeling with **CLARITE** and **ClariTeRep**

*https://github.com/jdaron/CLARI-TE*



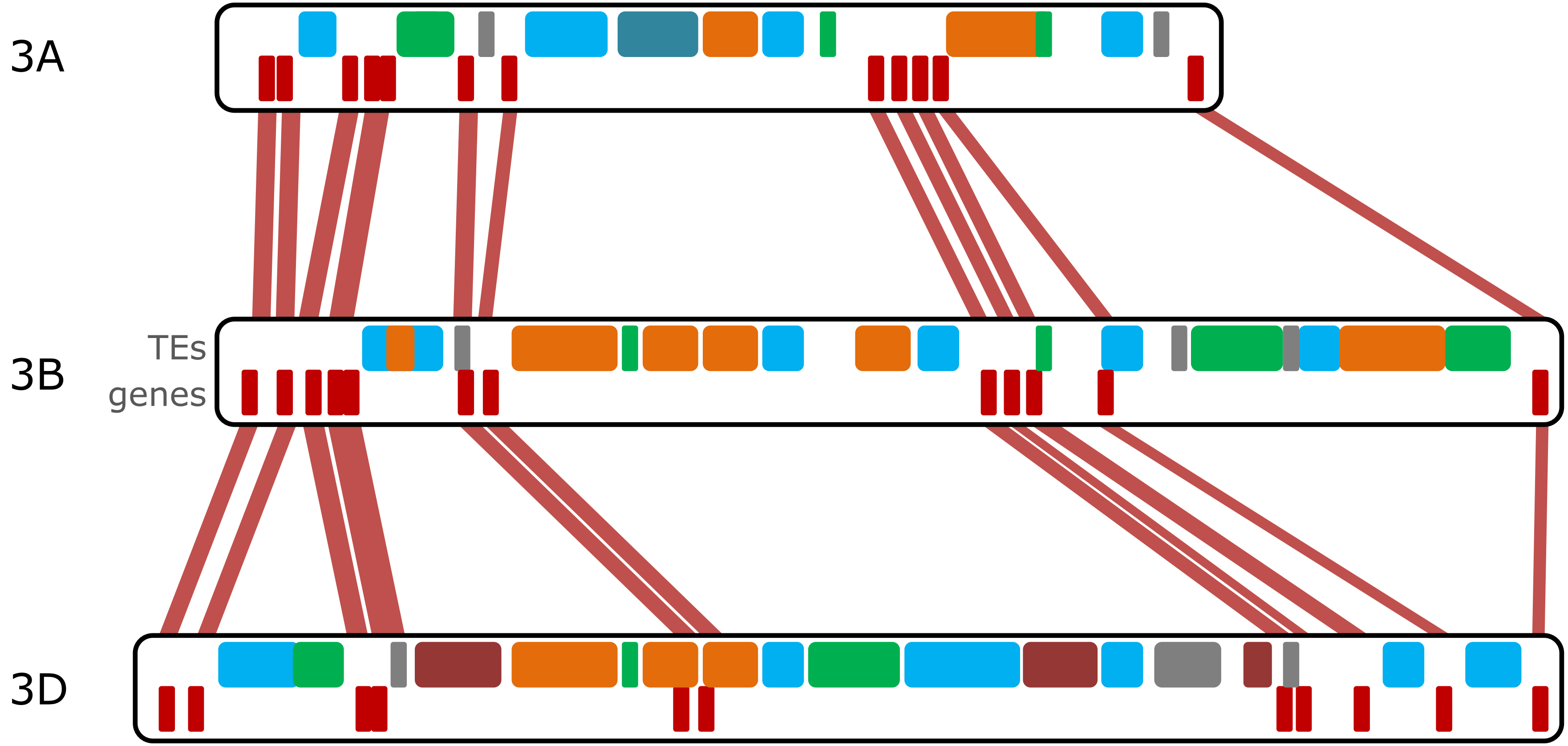


TREP++
5000 known wheat TEs

Classification

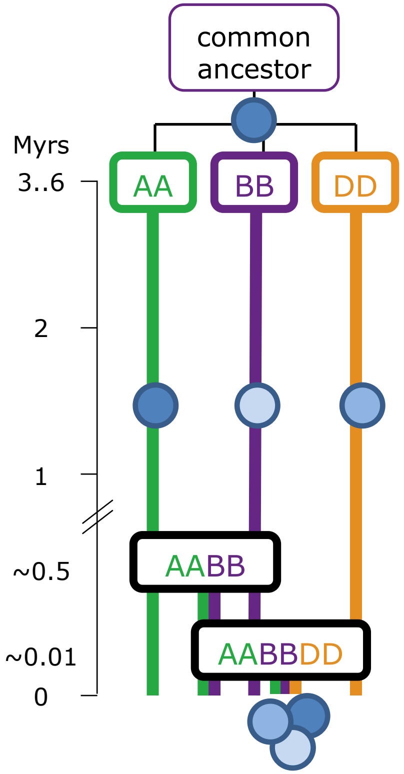
**ClariTeRep**
~500 families

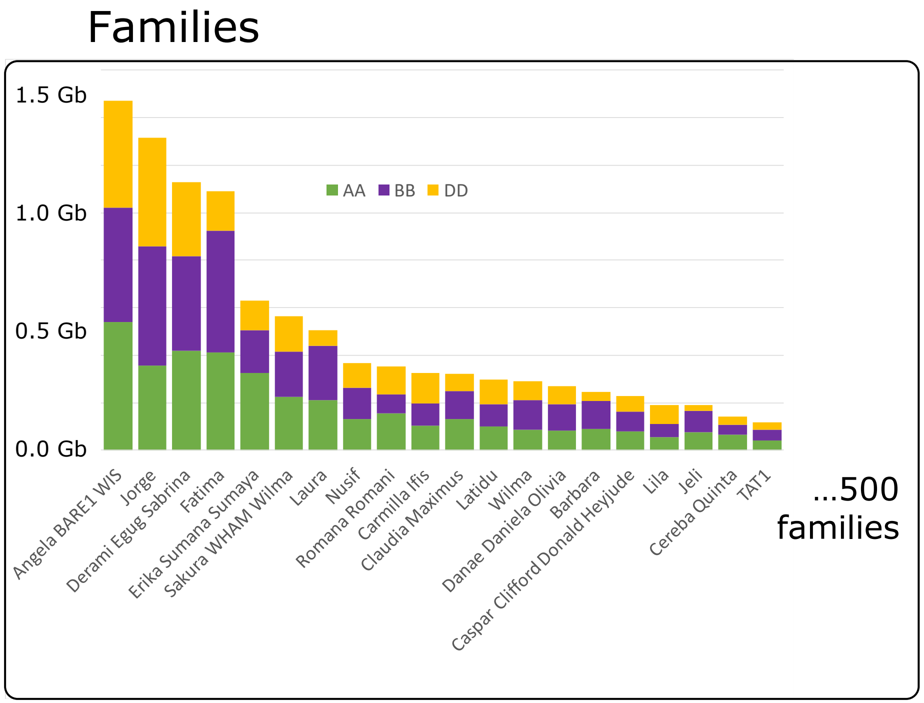
*Daron et al. Genom Biol 2015*



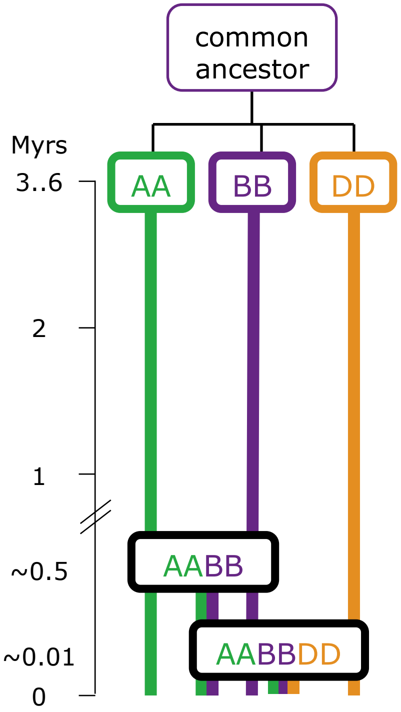
- Near-complete **TE turnover** since A-B-D divergence

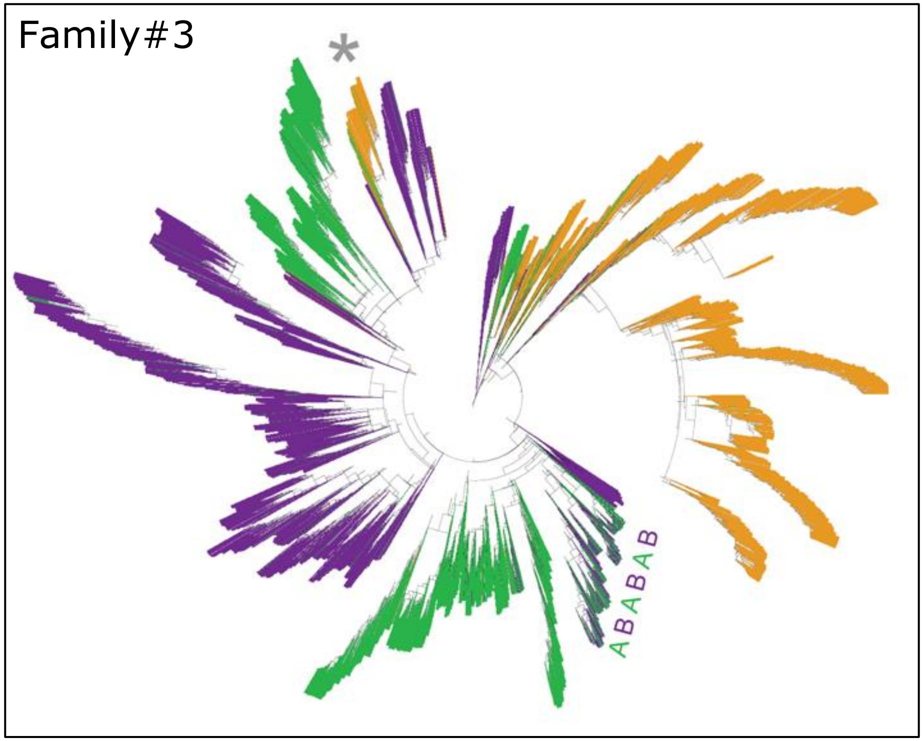
*Wicker et al. Genom Biol 2018*



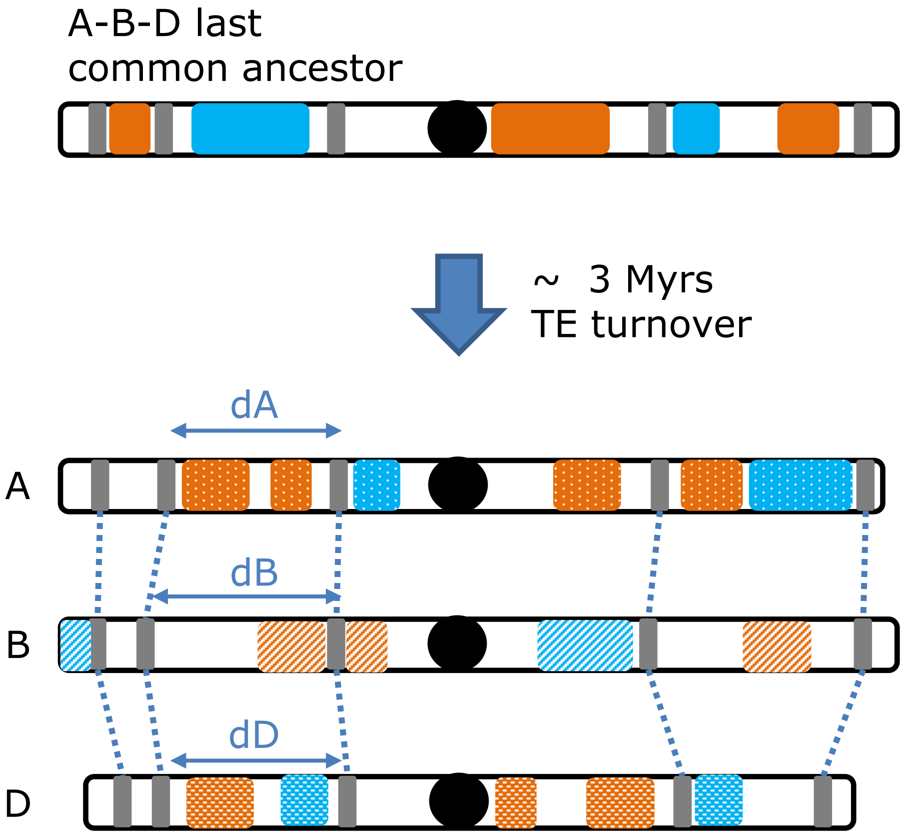


- A-B-D balanced proportions
- **0** family specific to 1 subgenome





- ➢ Independent TE evolution in the diploids AA, BB, DD
- ➢ Some rare cases of TE amplif in the tetraploid AABB



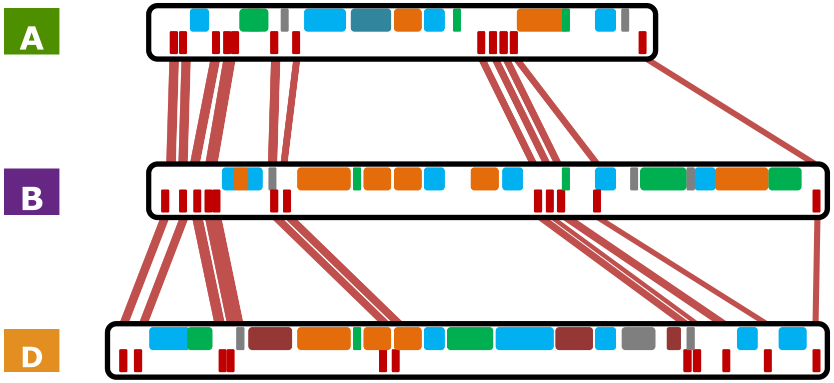
Most families were active since A-B-D divergence

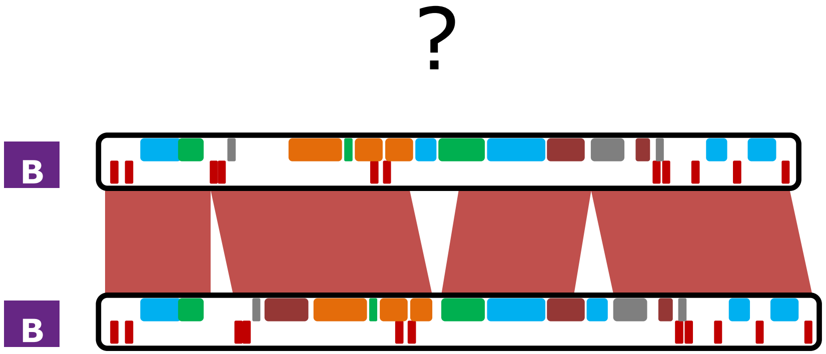
- ➢ **Equilibrium**

Hypothesis of a structural role of TEs likely under selection pressure?

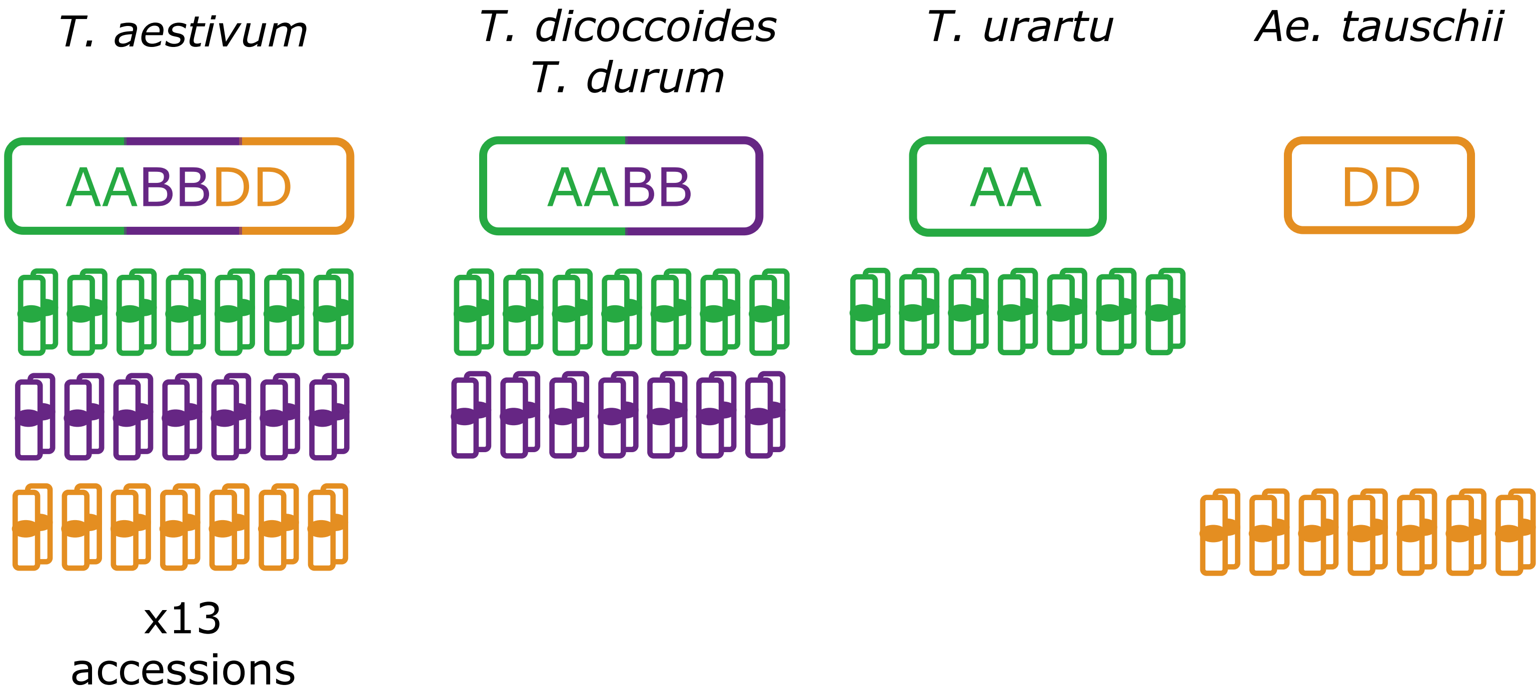
## ❑ Genomic variability affecting **TEs** in wheat **using assembled genomes**

*PhD Nathan Papon (2019-2022)*





# ○ Material = multiple RefSeq now available



*T. aestivum*

AABBDD

x13 accessions

*T. dicoccoides*
*T. durum*

AABB

*T. urartu*

AA

*Ae. tauschii*

DD

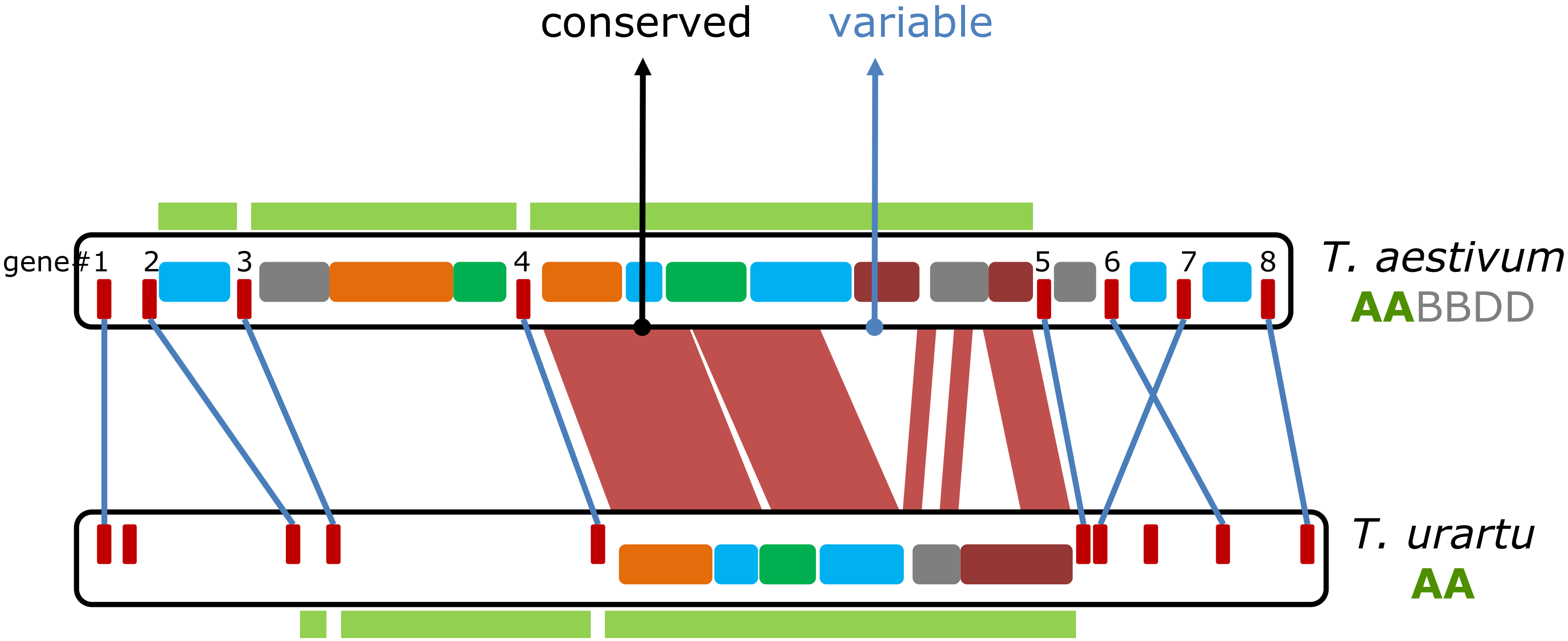
*IWGSC 2018*
*Walkowiak et al. 2020*
*Guo et al. 2020*
*Aury et al. 2022*

*Avni et al. 2017*
*Maccaferri et al., 2019*

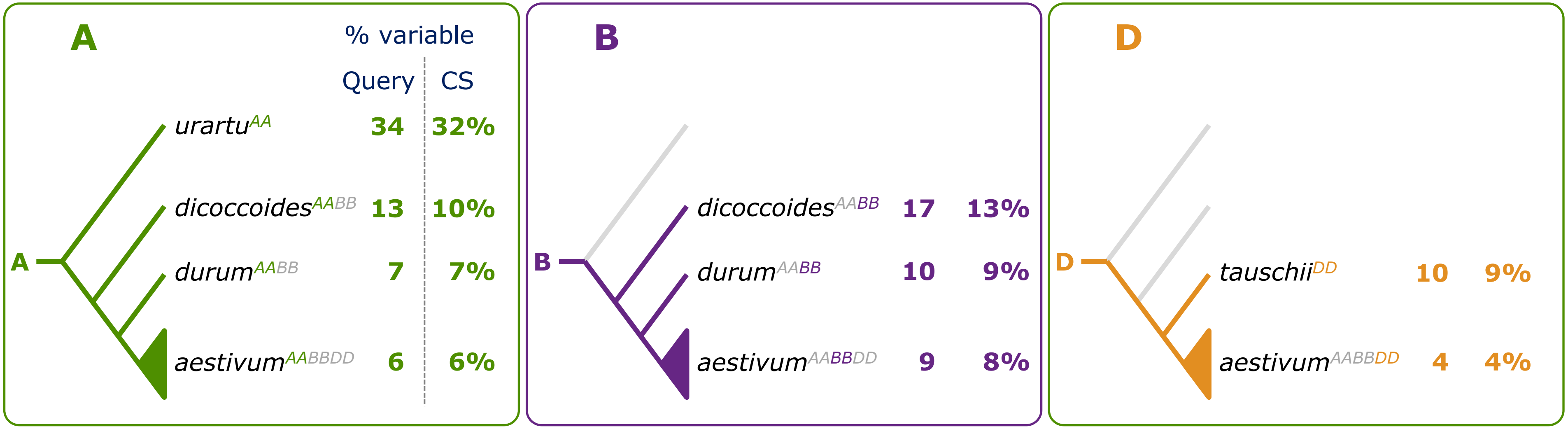
*Ling et al. 2018*

*Luo et al. 2017*

# ○ Method


conserved
variable
gene#1
2
3
4
5
6
7
8
T. aestivum
AABBDD
T. urartu
AA

# ○ Extent of variability affecting TEs



**A**

| | Query | CS |
|---|---|---|
| *urartu*$^{AA}$ | 34 | 32% |
| *dicoccoides*$^{AABB}$ | 13 | 10% |
| *durum*$^{AABB}$ | 7 | 7% |
| *aestivum*$^{AABBDD}$ | 6 | 6% |

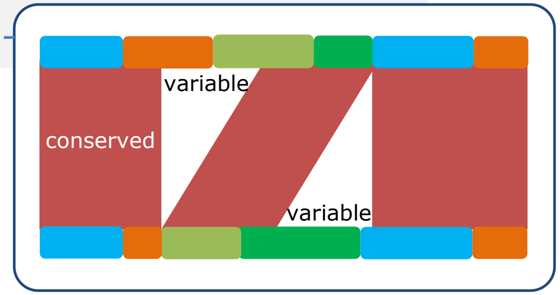
**B**

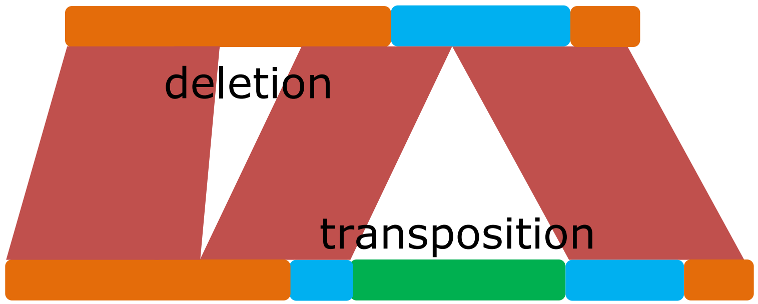
| | Query | CS |
|---|---|---|
| *dicoccoides*$^{AABB}$ | 17 | 13% |
| *durum*$^{AABB}$ | 10 | 9% |
| *aestivum*$^{AABBDD}$ | 9 | 8% |

**D**

| | Query | CS |
|---|---|---|
| *tauschii*$^{DD}$ | 10 | 9% |
| *aestivum*$^{AABBDD}$ | 4 | 4% |

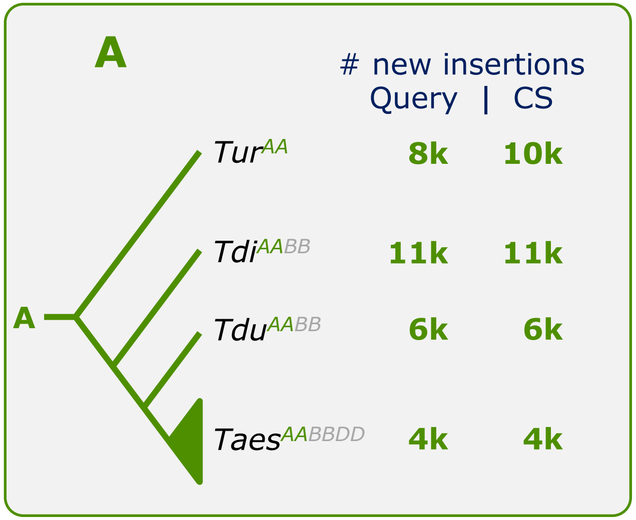
- 5% .. 15%
- B > A > D
- No genome expansion/contraction



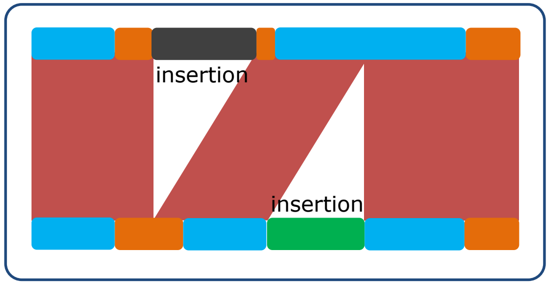


Recent TE insertions

- ~4,000–13,000 transposition events / subgenome (<1% of the genome)



| | # new insertions<br>Query | CS |
|---|---|---|
| *Tur*$^{AA}$ | 8k | 10k |
| *Tdi*$^{AABB}$ | 11k | 11k |
| *Tdu*$^{AABB}$ | 6k | 6k |
| *Taes*$^{AABBDD}$ | 4k | 4k |



Impact of polyploidy?

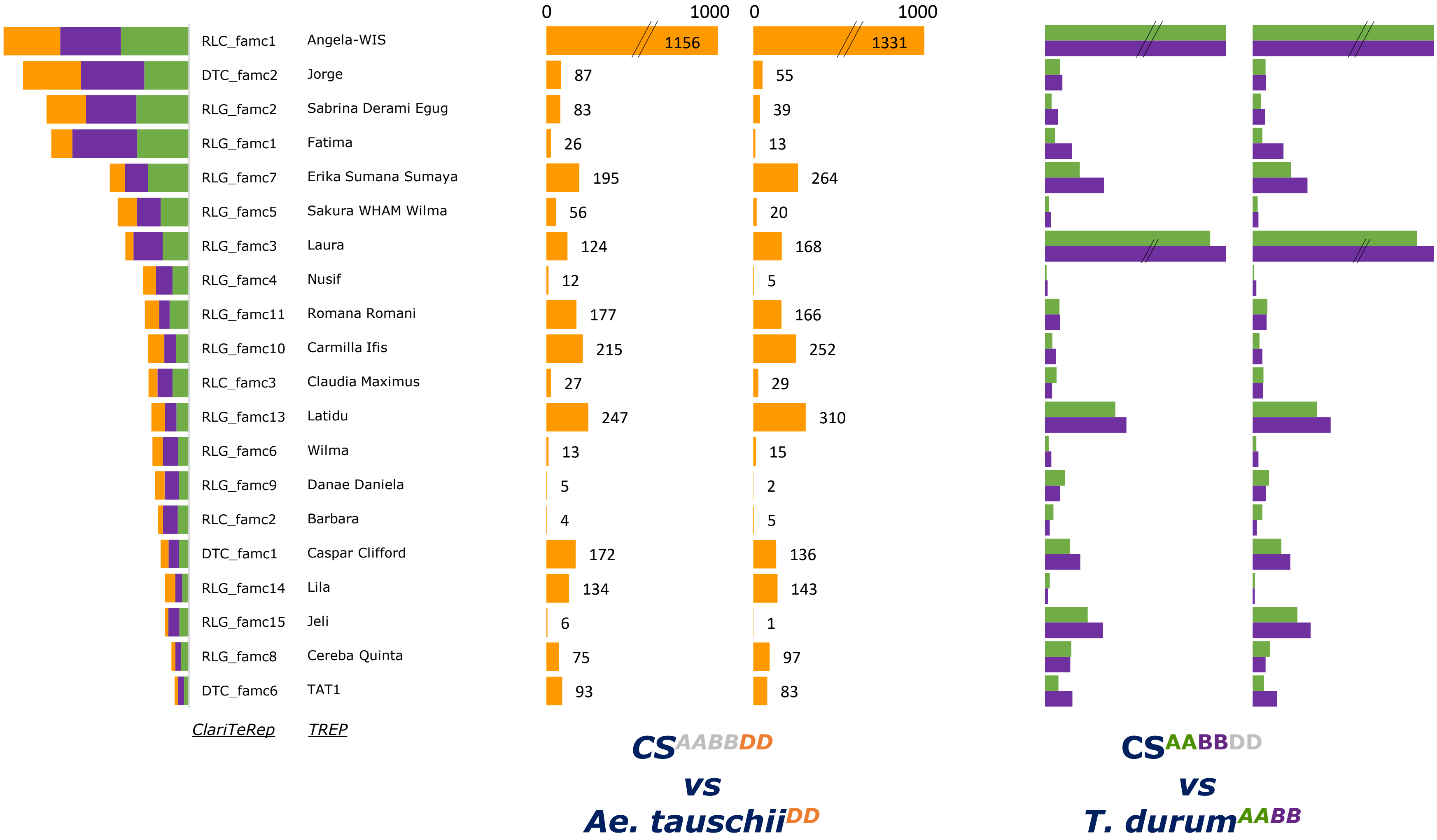
- none

How many families were active recently?

- 346 active families (=99.7% of all TEs)

# 20 most abundant families

# #specific TE insertions



# Conclusions

- No genome expansion/contraction due to TE transposition
- No impact of polyploidy
- No TE burst
- All families were active recently
- Transposition rate is stable

➢ Equilibrium

➢ Role of TEs in genome architecture under evolutionary constraints

# ❑ Acknowledgments

○ *Génoscope, CEA-CNRS-U Evry*
Jean-Marc Aury
Stefan Engelen
Benjamin Istace
Caroline Belser
Corinne Cruaud
Adriana Alberti
Patrick Wincker

○ *CNRGV, INRAE Toulouse*
Sandrine Arribat
Isabelle Dufau
Arnaud Bellec

○ *URGI, INRAE Versailles*
Michael Alaux
Nicolas Francillonne

○ *Mésocentre, UCA*
David Grimbichler

○ *GDEC, INRAE-UCA*
Cécile Monat
Pauline Lasserre-Zuber
Hélène Rimbert
Philippe Leroy
Etienne Paux
Marion Ranoux
Jérôme Salse
Carline Pont
Cécile Huneau
Nathan Papon
Romain De Oliveira
Josquin Daron

*Contact:*
*frederic.choulet@inrae.fr*

○ *Funding*

FRANCE GENOMIQUE

SRESRI/Région ARA

L'EUROPE S'ENGAGE en région Auvergne-Rhône-Alpes avec le FEDER

UNION EUROPÉENNE

La Région Auvergne-Rhône-Alpes