



Sequencing the genome of the French wheat variety Renan

Frédéric Choulet

► To cite this version:

Frédéric Choulet. Sequencing the genome of the French wheat variety Renan. IWGSC Webinar, IWGSC - International Wheat Genome Sequencing Consortium, Sep 2022, Clermont-Ferrand, France. hal-04090088

HAL Id: hal-04090088

<https://hal.inrae.fr/hal-04090088>

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.



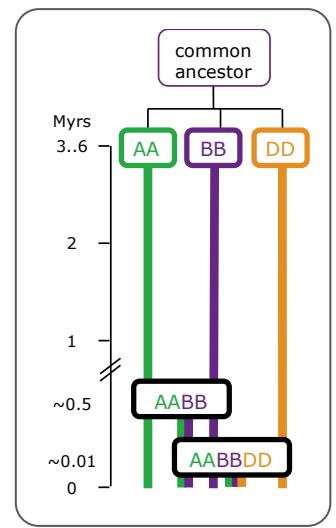
Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Sequencing the genome of the French wheat variety Renan

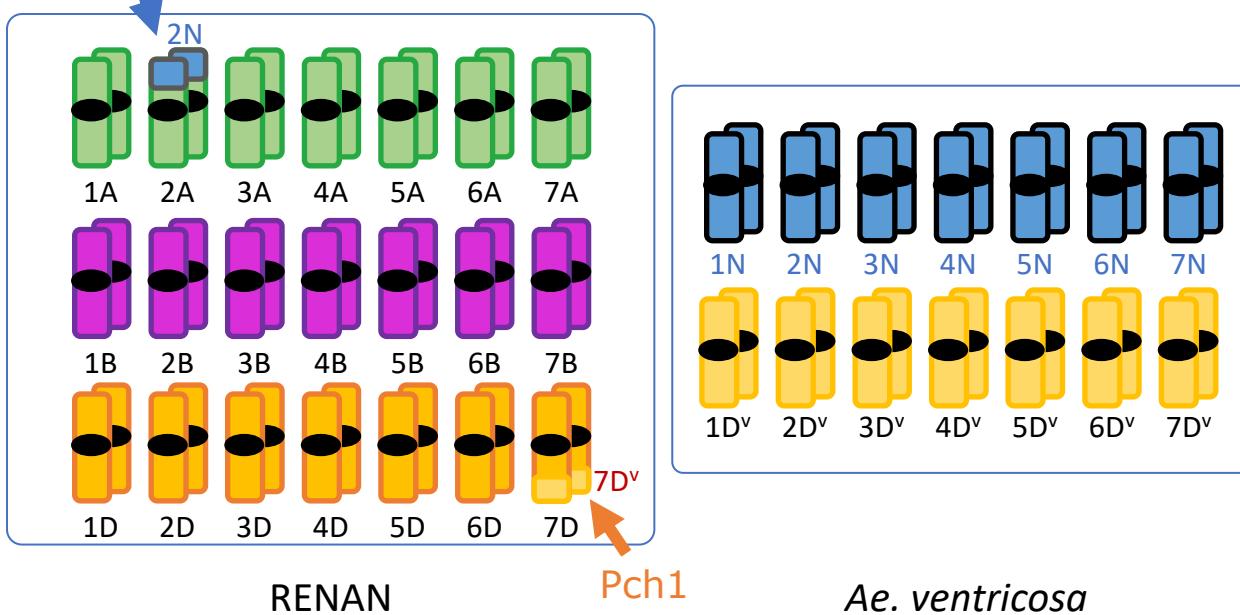
- *Frédéric CHOULET
GDEC, INRAE-Univ Clermont-Ferrand*

INRAe

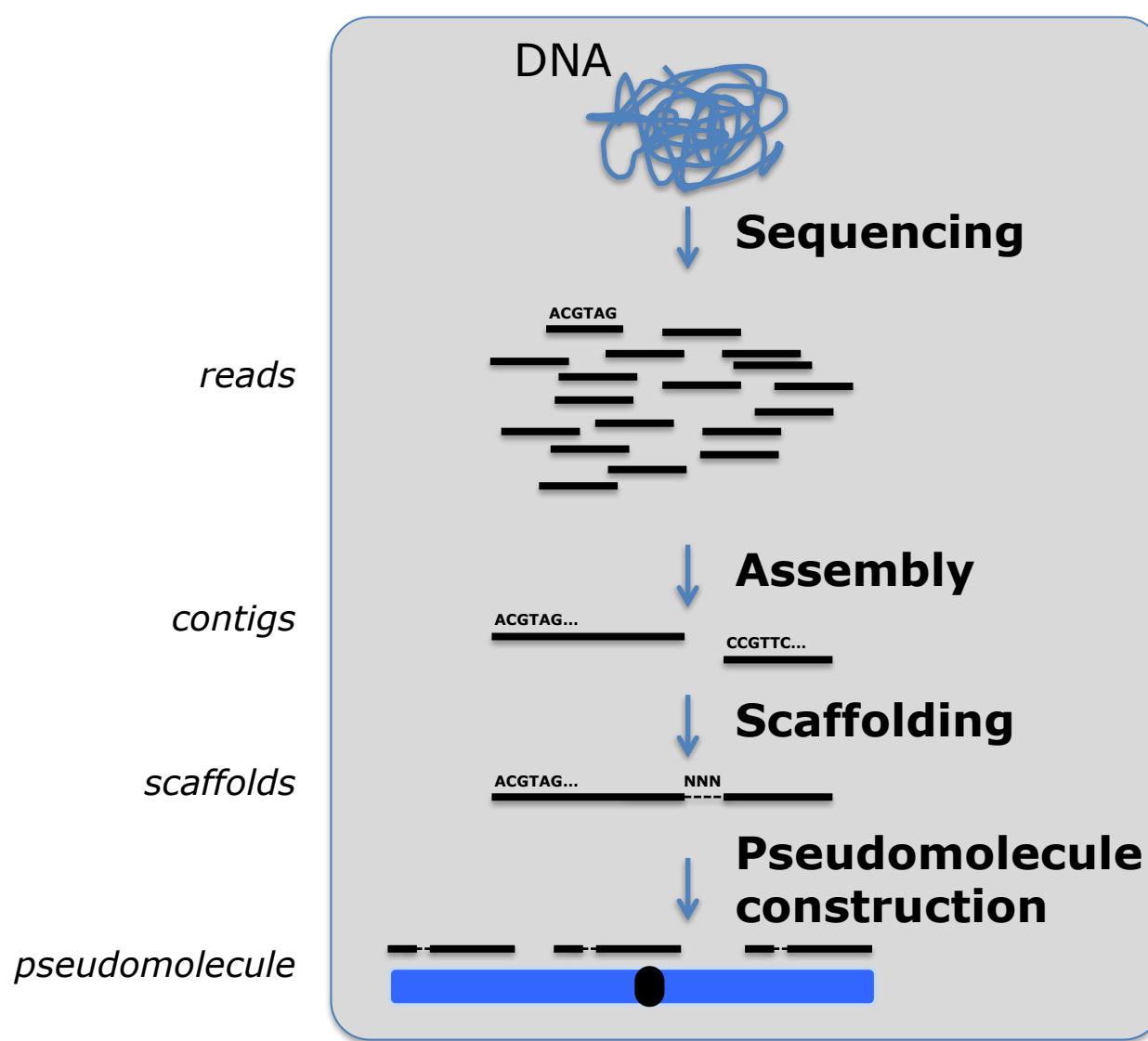

UNIVERSITÉ
Clermont
Auvergne



Yr17 Lr37 Sr38 Cre5



□ Genome assembly: steps and technos



Illumina



Pacbio
Nanopore



NRGene
(*DeNovoMAGIC*)



Bionano



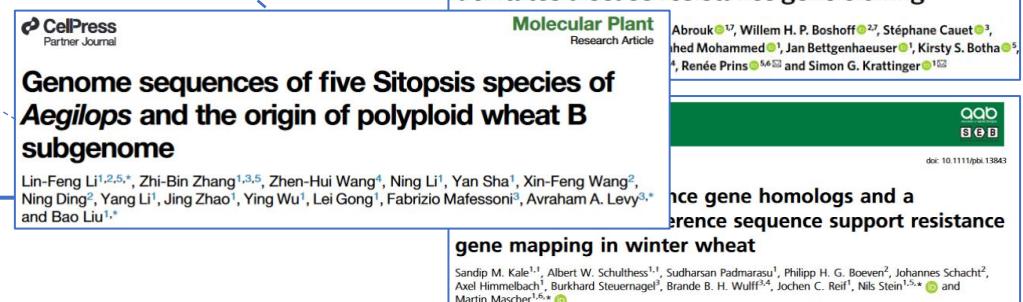
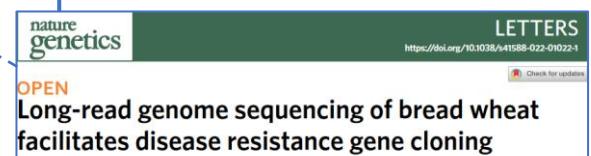
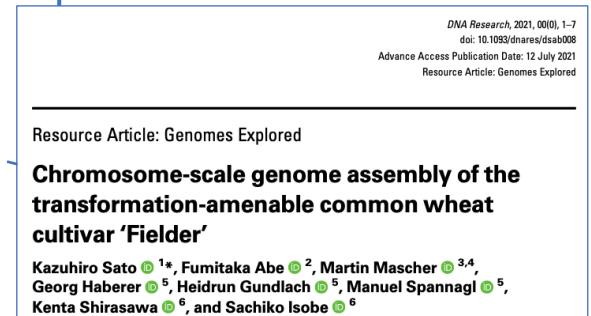
HiC



☐ *Triticeae* refseq current status (09/2022)

Reference-quality genome assemblies

- *T. aestivum*
 - Chinese Spring (NRgene)
 - ArinaLrFor 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



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, 2}, Cécile Monat^②, Pauline Lasserre-Zuber², Caroline Belser^{①, 2}, Corinne Cruaud^③, Hélène Rimbert^②, Philippe Leroy^②, Sandrine Arribat^④, Isabelle Dufau⁴, Arnaud Bellec^④, David Grimbichler^⑤, Nathan Papon², Etienne Paux^②, Marion Ranoux², Adriana Alberti^{①, 6}, Patrick Wincker^① and Frédéric Choulet^{②, *}

¹Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

²GDEC, Université Clermont Auvergne, INRAE, UMR1095, 63000 Clermont-Ferrand, France

³Commissariat à l'Energie Atomique (CEA), Institut François Jacob, Genoscope, F-91057 Evry, France

⁴INRAE, CNRGV French Plant Genomic Resource Center, F-31320, Castanet Tolosan, France

⁵Mésocentre Clermont Auvergne, DOSI / Bâtiment Turing, 7 avenue Blaise Pascal, 63178 Aubière, France

⁶Present 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

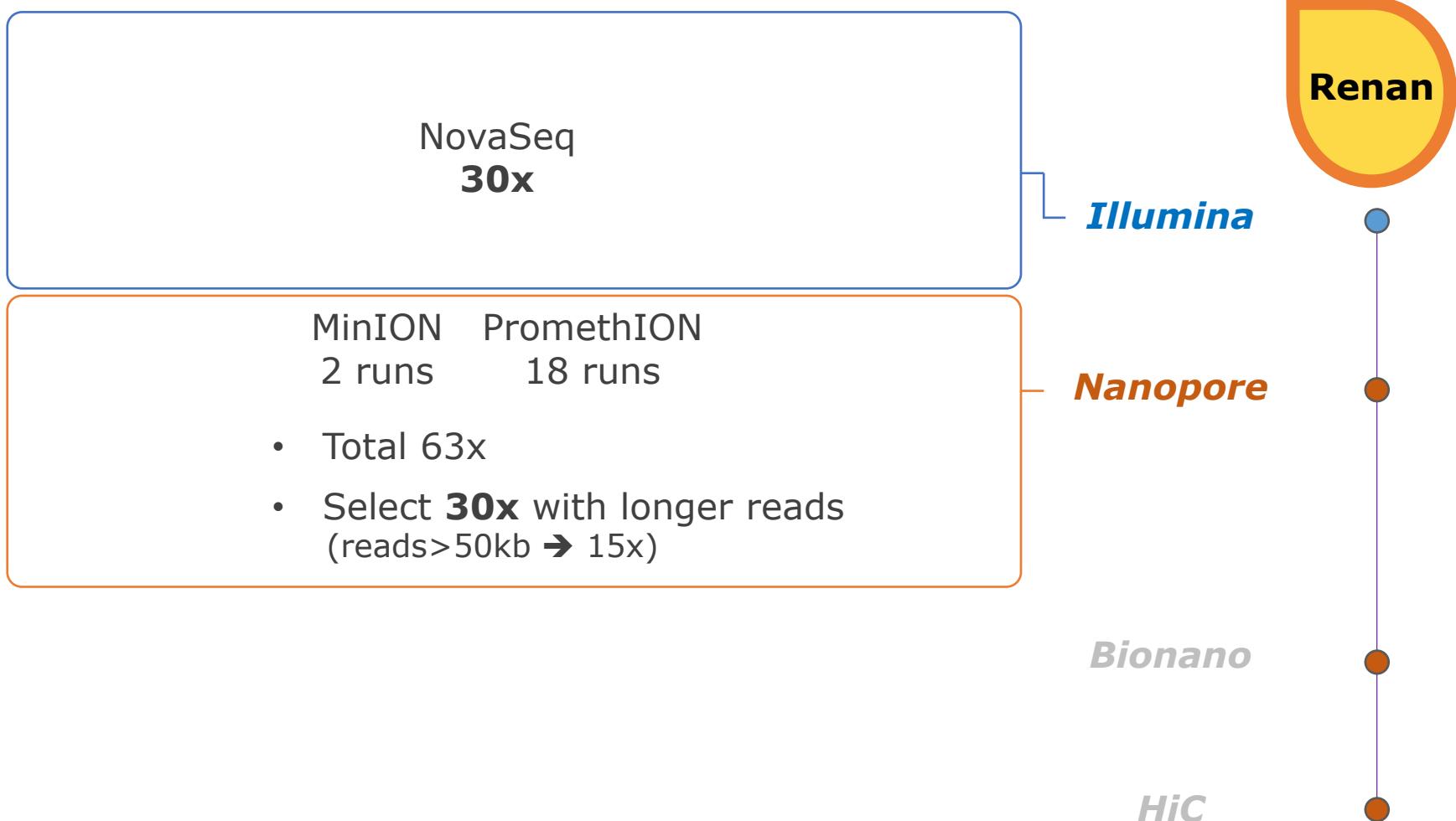
Background: The sequencing of the wheat (*Triticum aestivum*) genome has been a methodological challenge for many years owing to its large size (15.5 Gb), repeat content, and hexaploidy. Many initiatives aiming at obtaining a reference genome of cultivar Chinese Spring have been launched in the past 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.

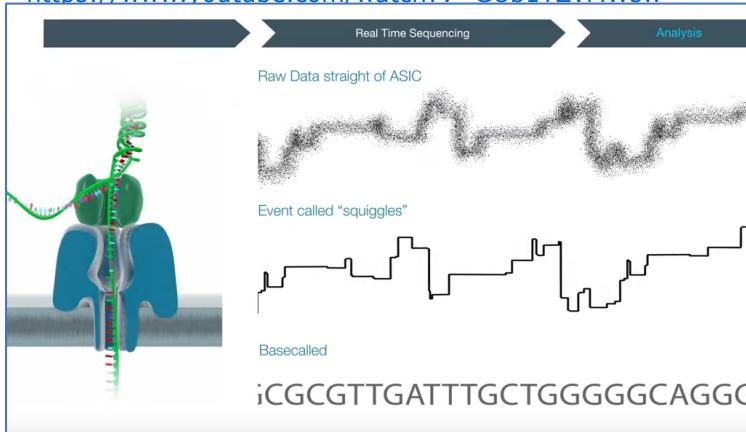
Keywords: wheat, hexaploid genome, long-reads, nanopore sequencing, genome assembly, haplotype characterization, introgressions

☐ ***Step#1: sequencing***



☐ Step#2: read assembly

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



Early stages

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

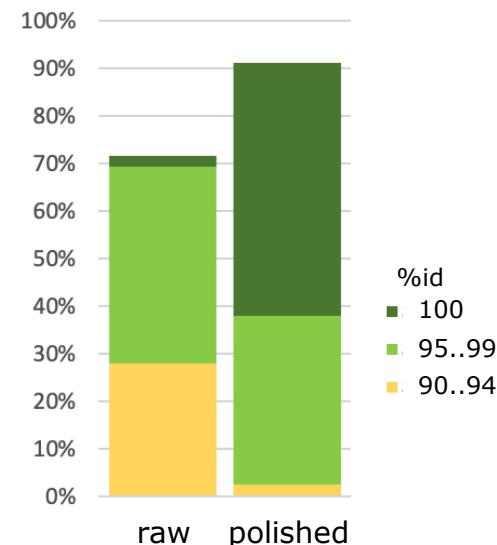
- basecalling
- polishing



ACGTAGACGTACGGCTGCA**T**CGCATTCATGGCGGGGGGG**G**CTGCTCGTCGCTCCTCGT

assembly

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



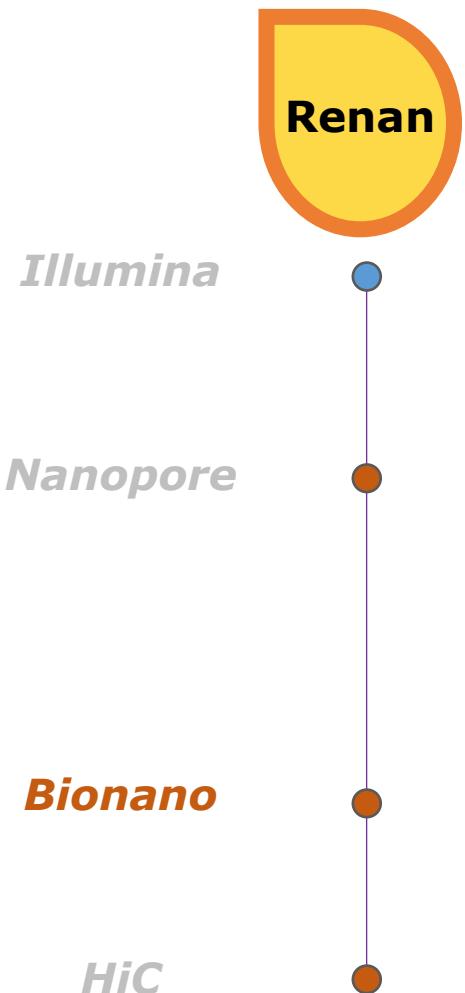
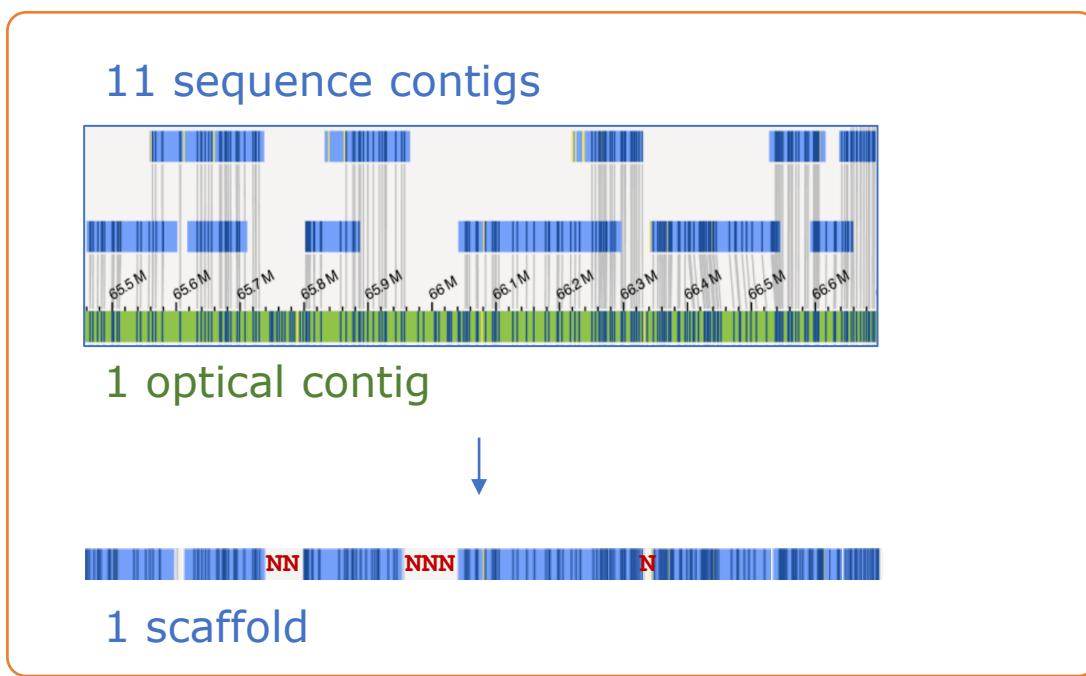
Renan

Nanopore



Illumina
30x

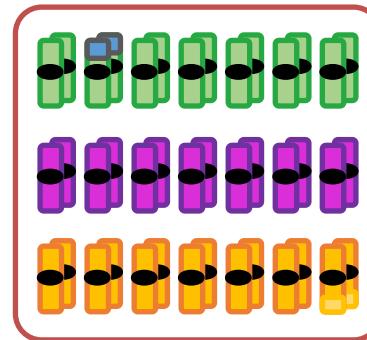
□ ***Step#3: scaffolding***



□ ***Assembly metrics***

❖ **Metrics**

- Genome size=14.3 Gb
- 2904 scaffolds (=138/chr)



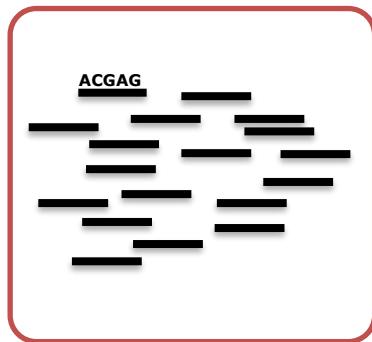
❖ **Quality Assessment**

Genes: 98% CS genes found

TEs: 92% CS ISBPs found

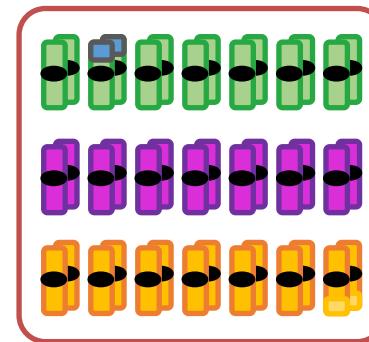
QV: lower (99.95% accuracy) than NRGene-based assemblies (99.99%)

□ ***Step#4: pseudomolecule construction***



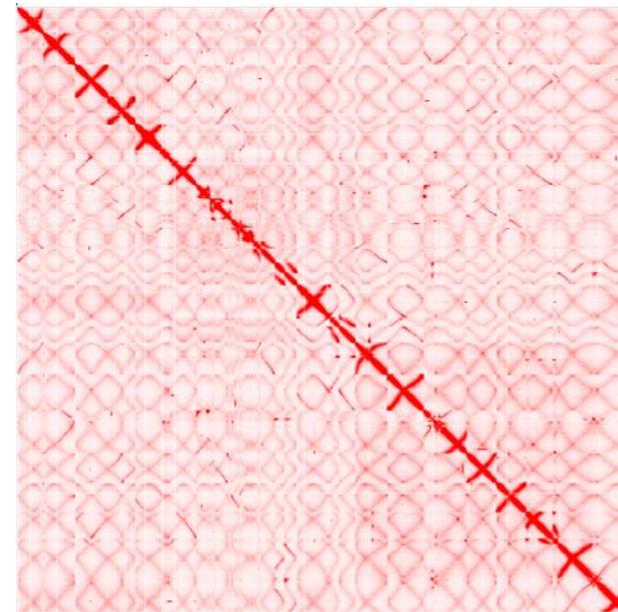
2904 scaff

1. CS-guided
2. HiC

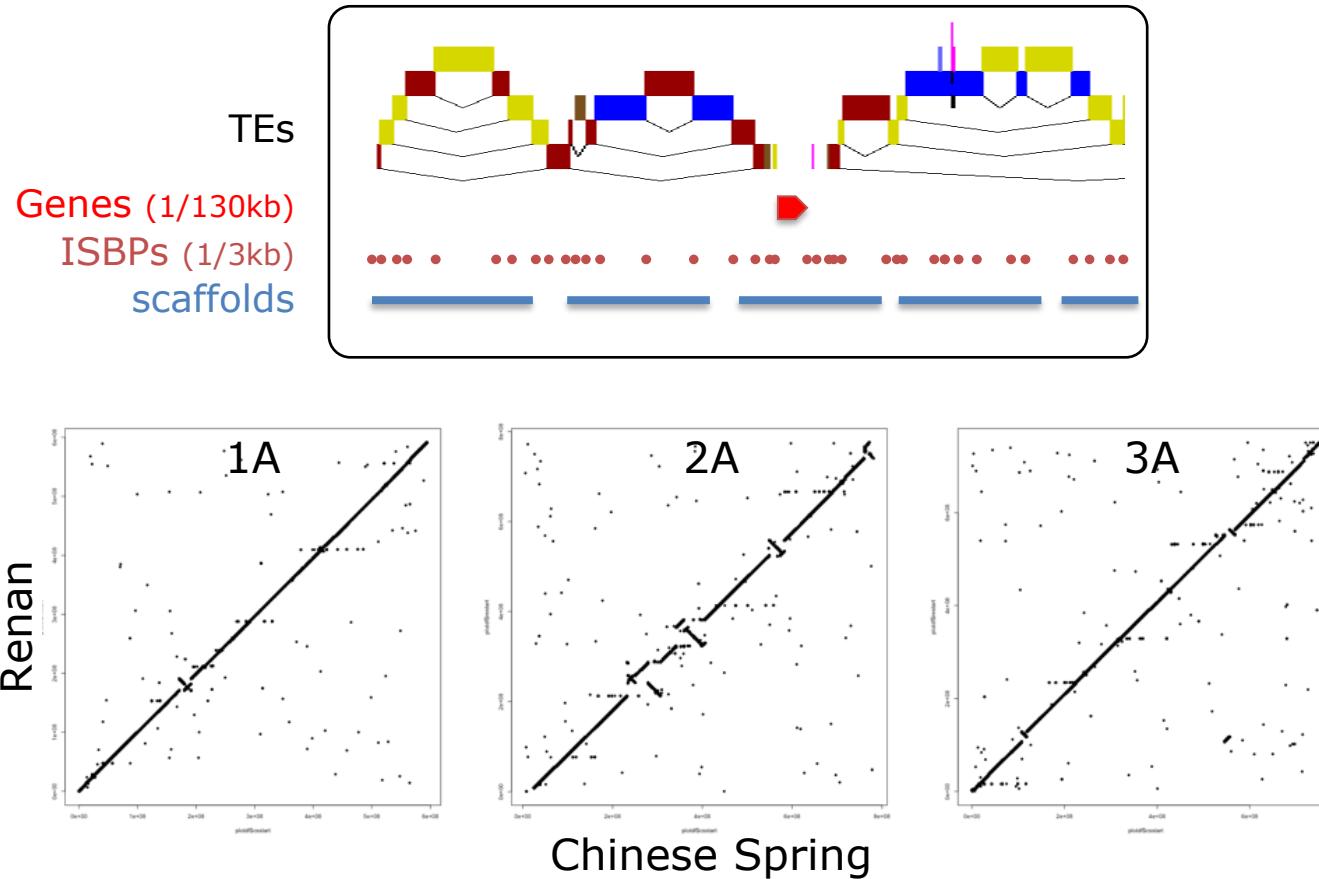


21 chromosomes

✗ not efficient enough to build a chromosome-scale assembly

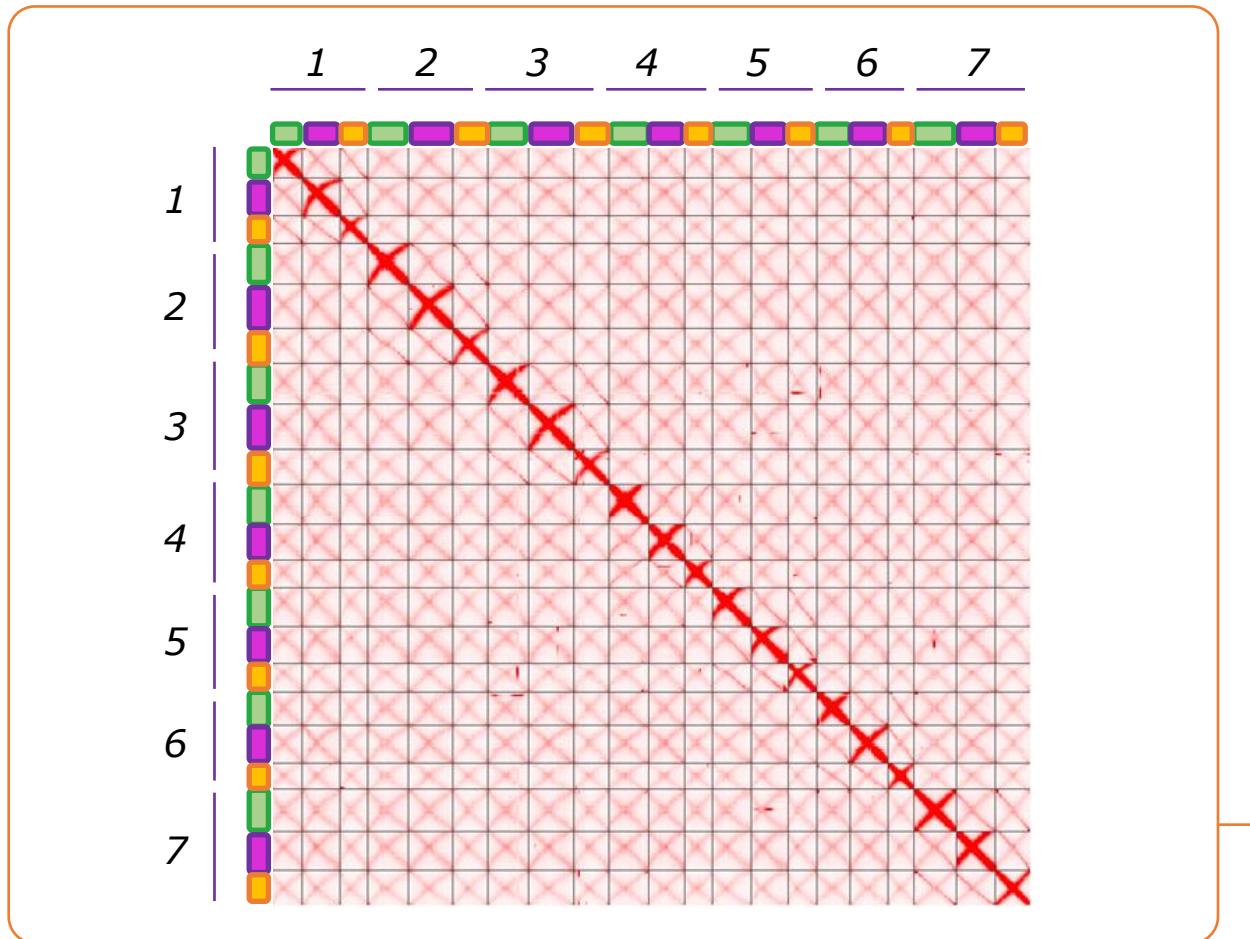


□ Step#4-1: CS-guided pseudomol construction



⇒ 14.2 Gb (0.1 Gb unanchored)

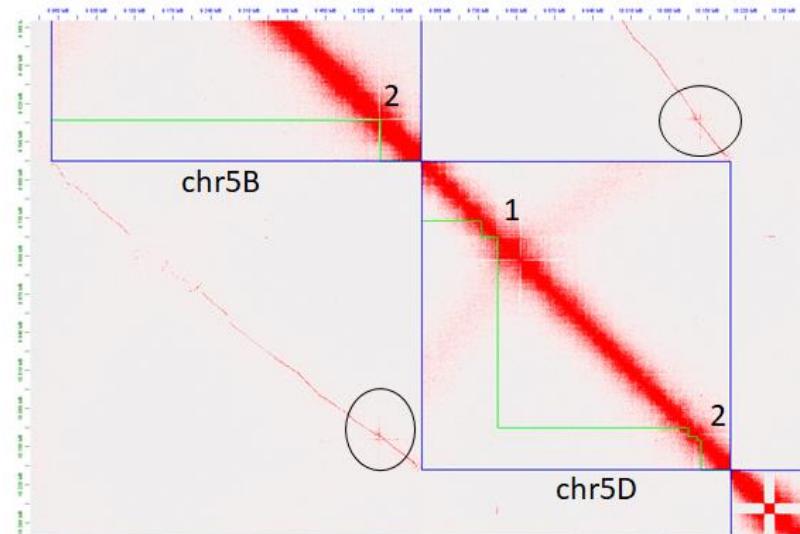
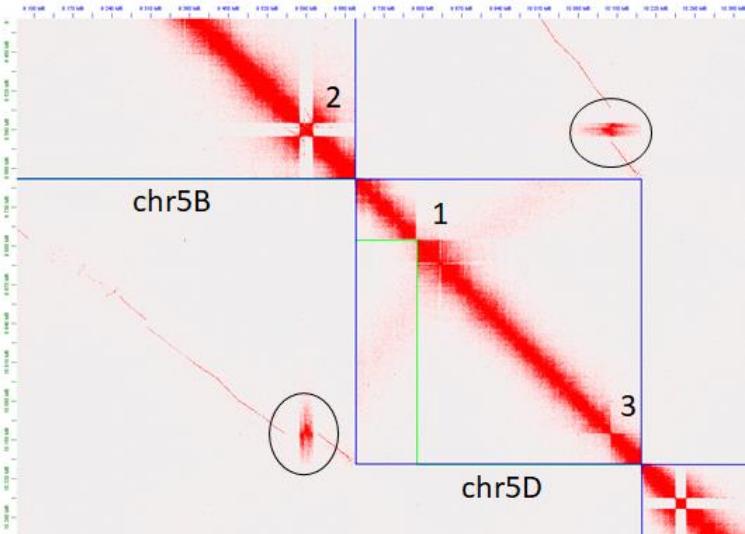
- *Step#4-1: CS-guided pseudomol construction*
- *Step#4-2: HiC corrections*



HiC



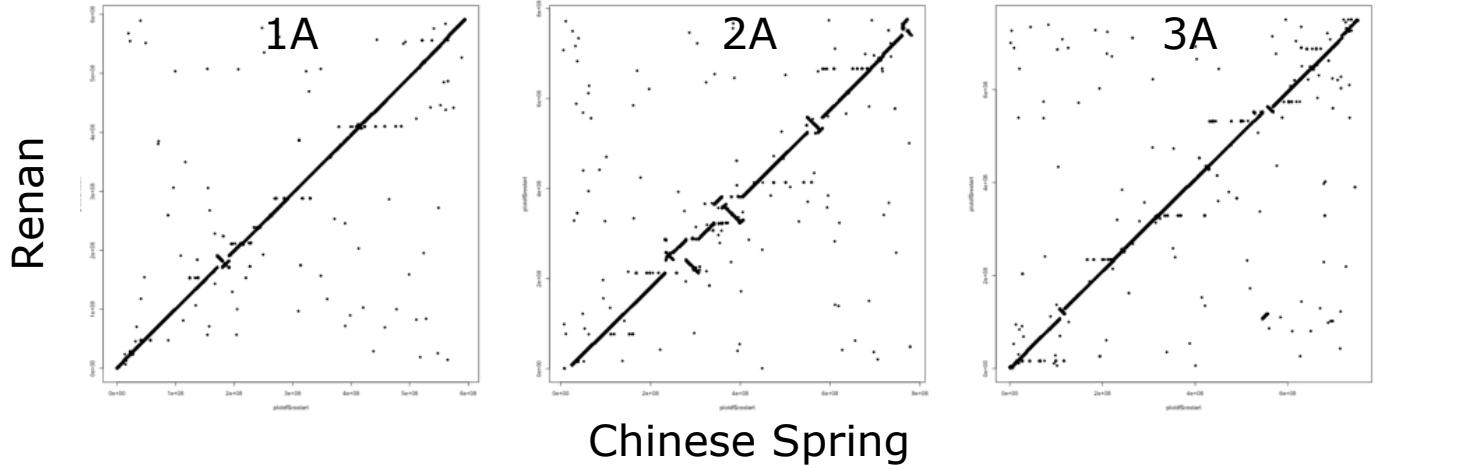
□ Step#4-2: HiC corrections



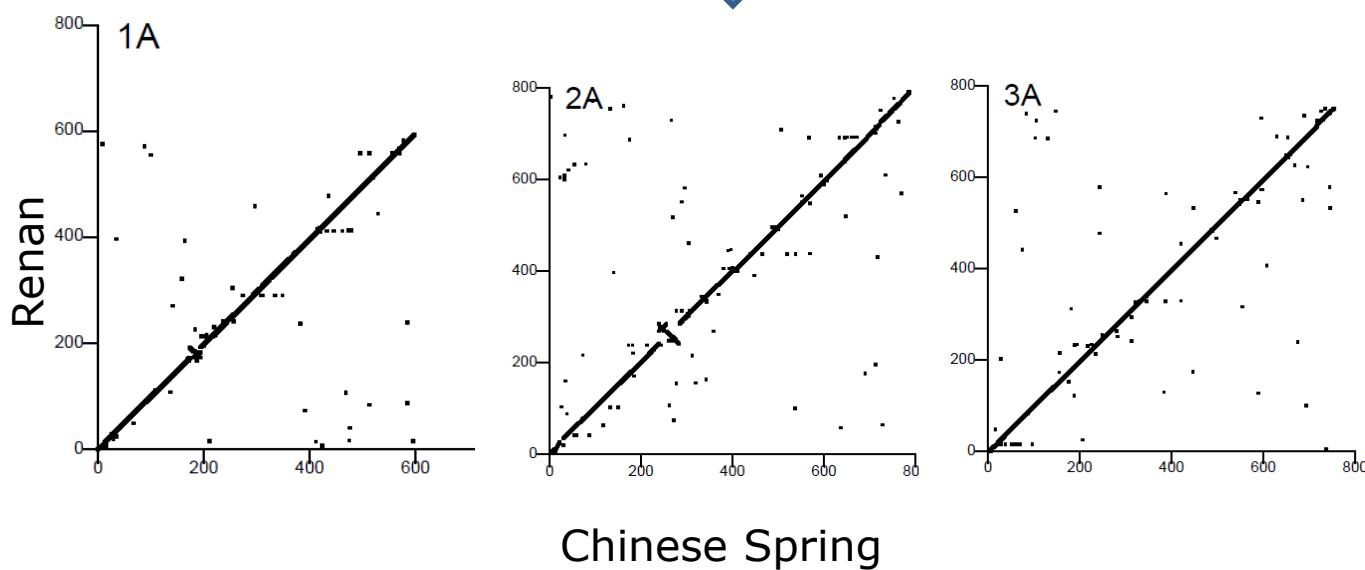
- **18** chimeric scaff
- **198** scaff with correction of location and/or orientation



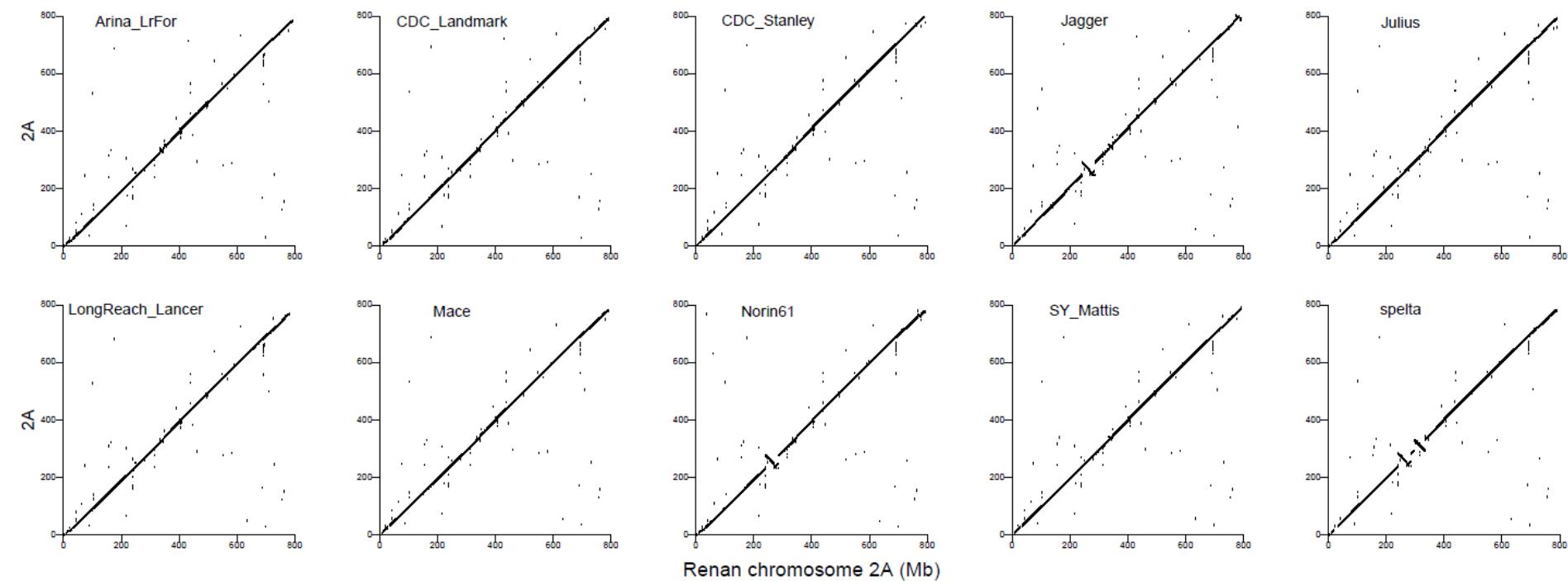
21 chr + chrUn 61 Mb (<1%)

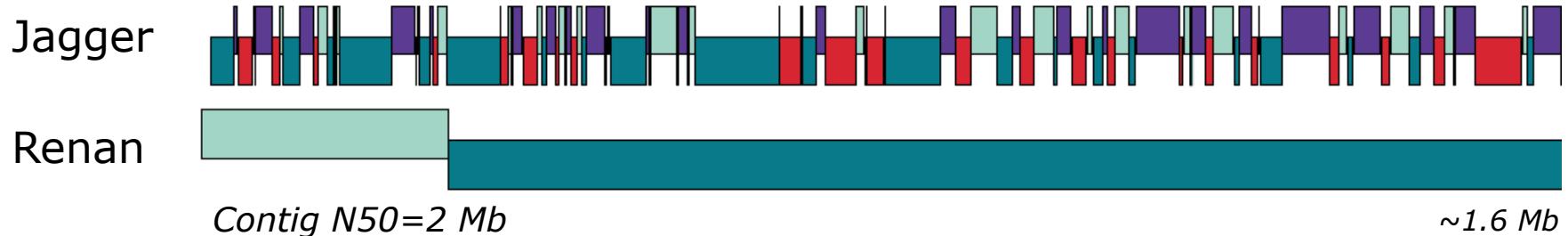


After HiC-based corrections

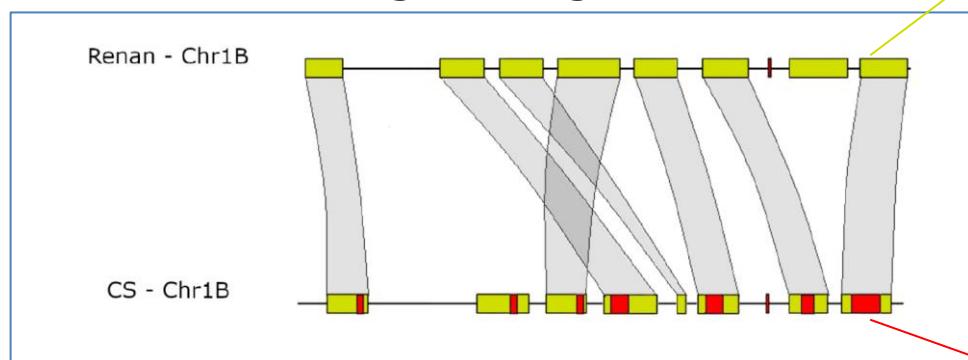


Renan Refseq_v2 versus 10 wheat genomes





ω -gliadin gene cluster



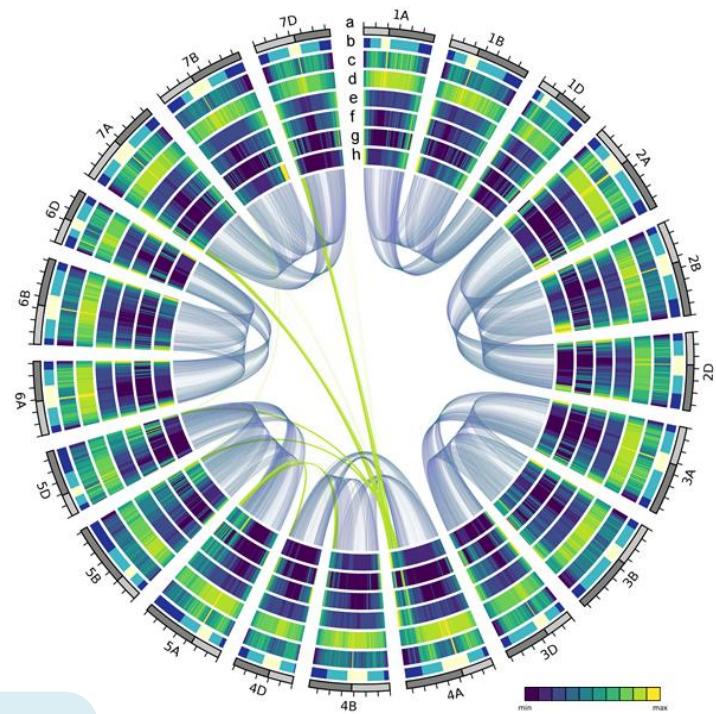
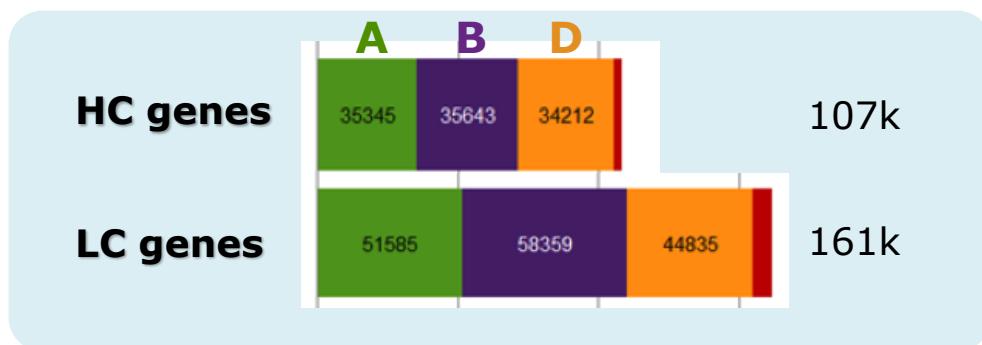
ATGAAGACCTTCATCATTTGTCCTCCTGGTATGGTATGAGCATGCCAGTGCCTAGGCAACTAA
GCCCTAGAACGAAAGGAGTTGCAACACCACAAGAACAAATTCCCCAACAGCAATTCCCTAACCAACAA
AATCTCTC**AAACAACAAATTCCCCAACACAAATCCCCAGCAACCCAAACAAATCCCCAACAAACAA**
ATCCCGAACAGAACAAATTCCCGAACACAAATCCCCAACACAAACAAATCCCCAGCAACCCAACAAAT
CCCCAACACAGAACAAATTCCCCAACACAAACAAATCCCCAGCAACCCAACAAACAAAT
CCCCAACACAGAACAAATTCCCCAACACAAACAAACAAATCCCCAGCAACCCAACAAACAAAT
GAATTCCCCAACACAAATCCTCCAGCAACCACAAACATTCCCCAACACAGAACATTCCCCAA
ACAATTCCCCAACACAAATCCCCGAAC // ACAATTCCCCAACAGCAATTCCCCAACACAAAC
AATTCCCCAACACAAACCCAGTTACCGAACACAAATCCCCAACACAAACAAATCTCCGAGCACACCA
ACAATTCCCCAACACAAACAAATCCCCAACACAAACAAATCCCCAACAGCAATTCCCCAACACAA
TTTCCCCAACACAAACAAATCCCCAACACAAACAGTTCCTCAACAAACAAACAAATCCCCAACACAA
TCCCTGAAACACAATAATTTCCCCAACACAAATCCCCAGCAACCATCACAAACAAATTCCCCAACACAA
ATTTCCCAATACCATACCCACCCCAACAAATCACAGAACACCTTCCCCATACCAACAAATATCCACAAACAA
CCATATGGGAGAACGCTTAAAGTATCTAGTCTGGCGGATGA

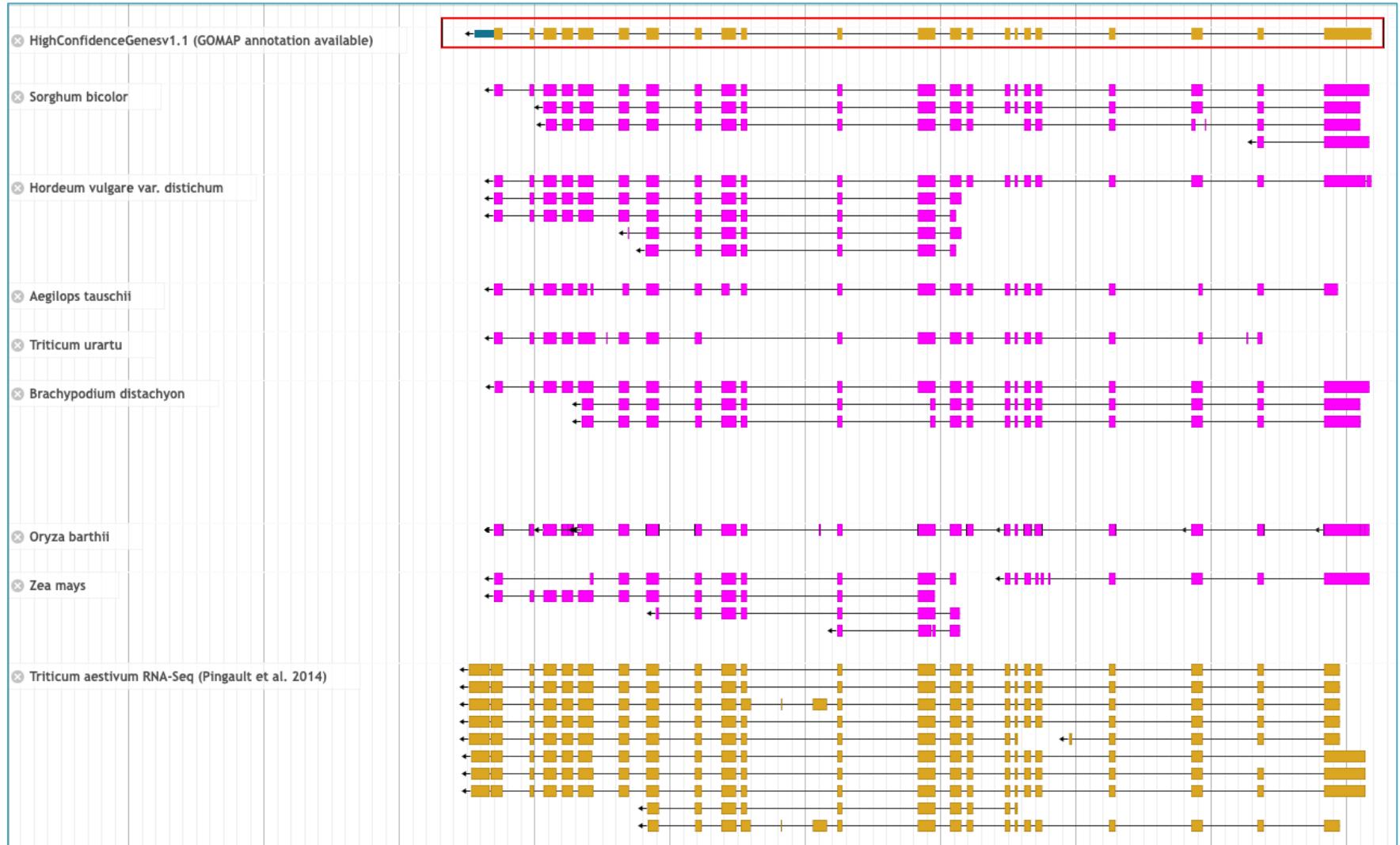
○ Predicting genes is still NOT routine

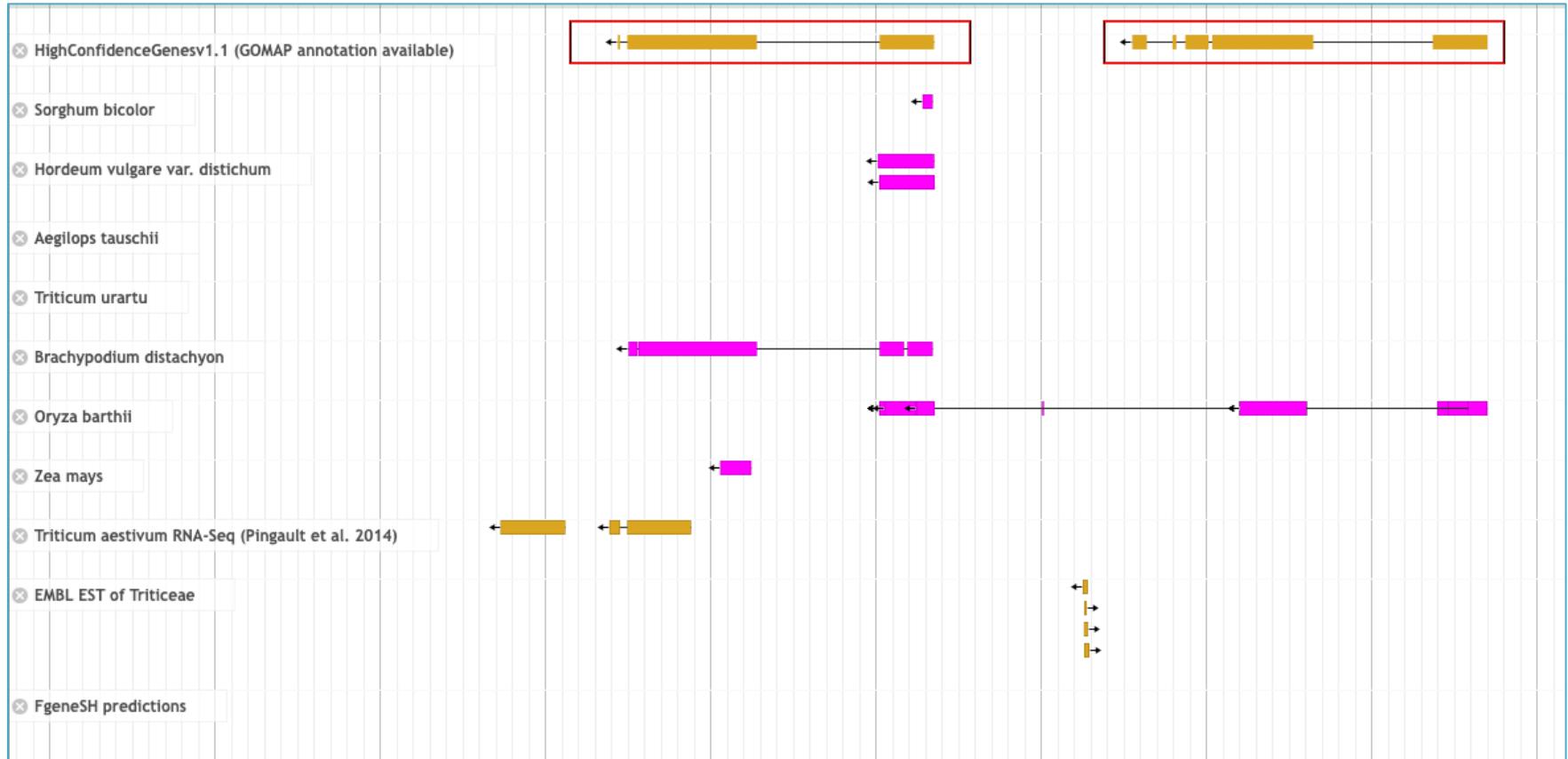
GDEC Clermont, France – *Rimbert Leroy Choulet et al.*
PGSB Munich, Germany – *Spannagl Twardziok et al.*
EI Norwich, UK – *Swarbreck Venturini et al.*

IWGSC RefSeq (Chinese Spring)

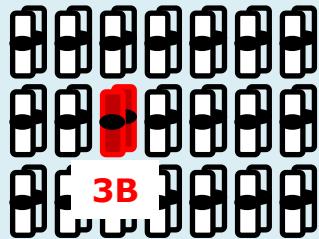
- 107,891 genes



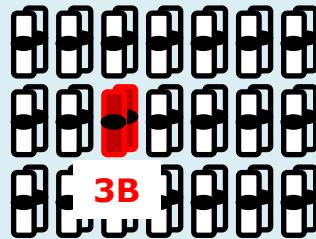




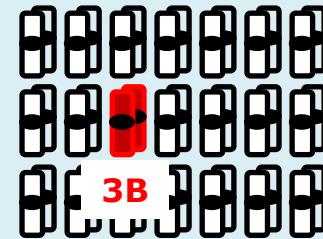
chr3B BAC-by-BAC



TGAC_v1



Chr. Survey Seq



7264 genes



5728 genes

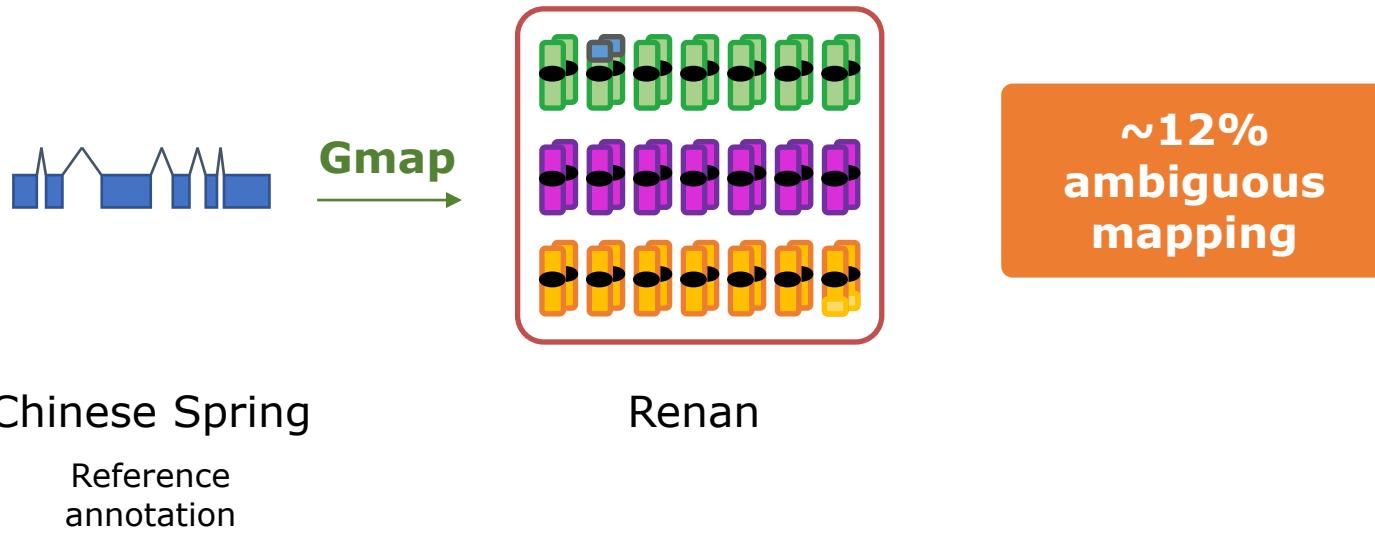
• (26% identical)

5862 genes

• (12% identical)

- *denovo* gene annotation not adapted to pangenomics

- Gene projections / Annotation transfer



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



magatt

Project ID: 1089

[Star](#) 1

• 120 Commits 1 Branch 5 Tags 5.5 MB Project Storage

Pipeline used to transfert 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 Singulair...	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.

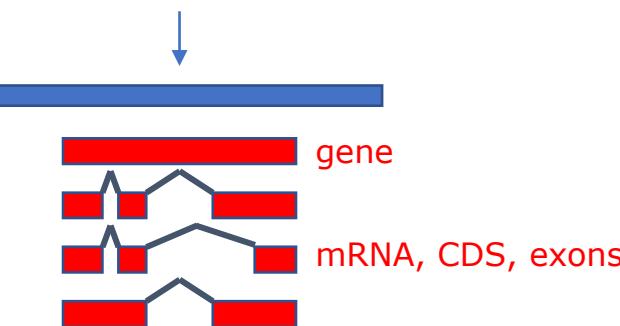
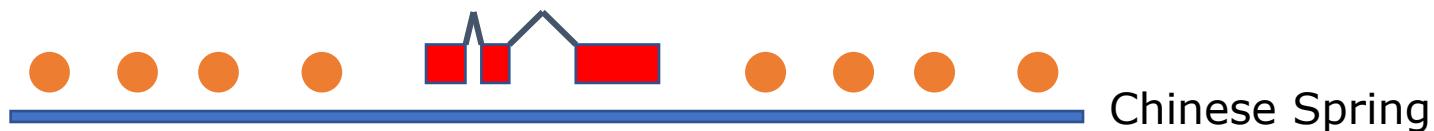
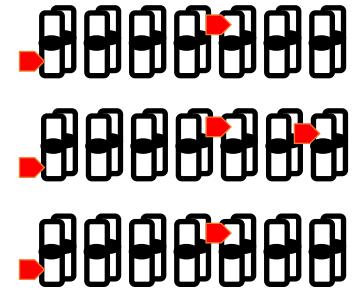
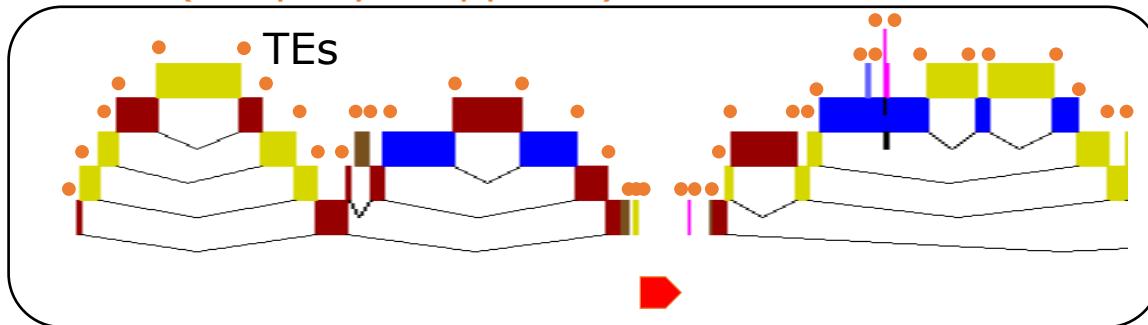
Snakemake pipeline used to transfert GFF annotation on a new assembly with a fine target mapping approach.

Install the pipeline

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

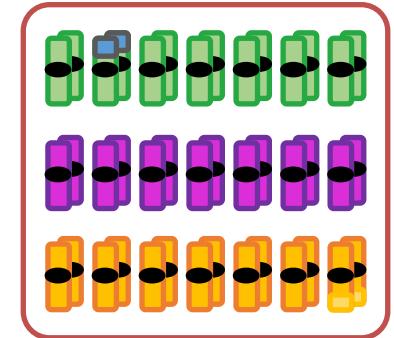
□ MAGATT (Marker-Assisted Gene Annotation Transfer for *Triticeae*)

ISBPs (uniquely mappable)



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

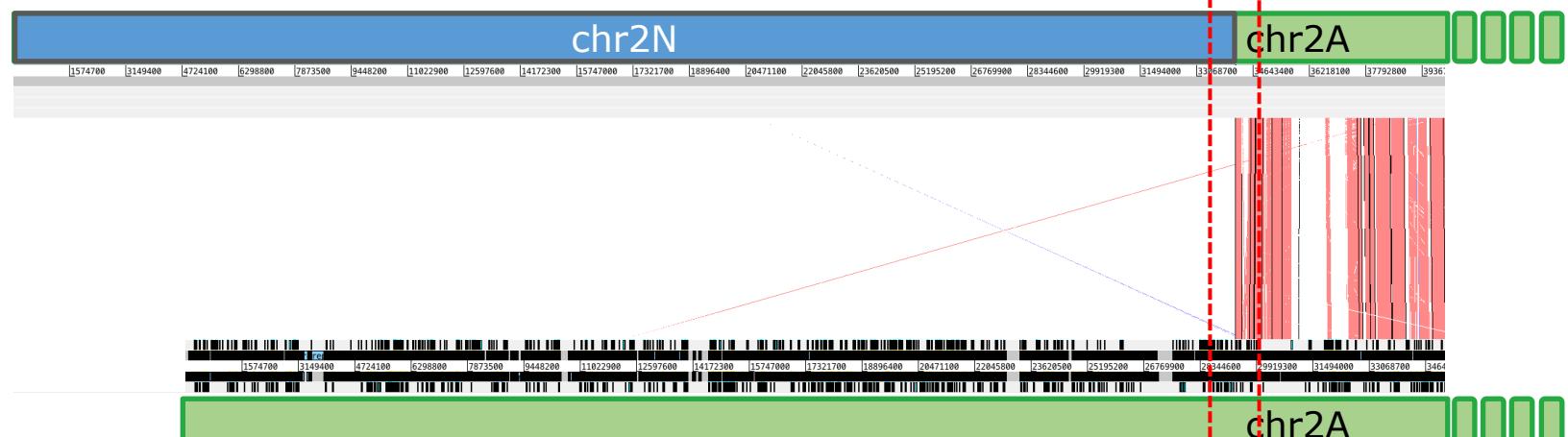
☐ *Introgressions*



telomere

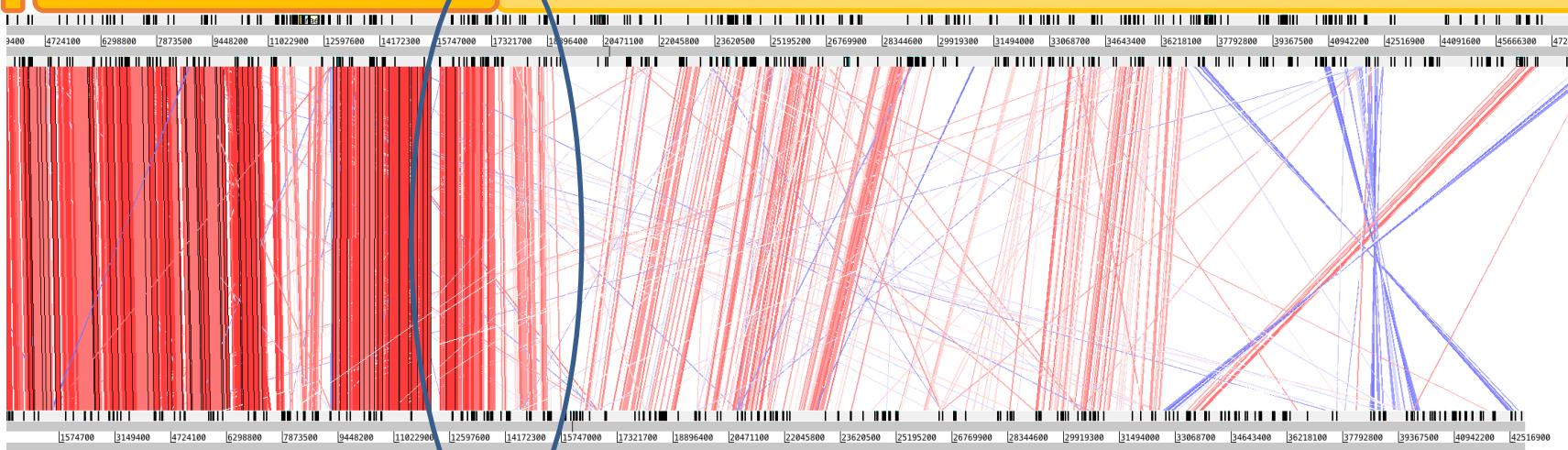
Renan

40 Mb



Renan – chr7Dv

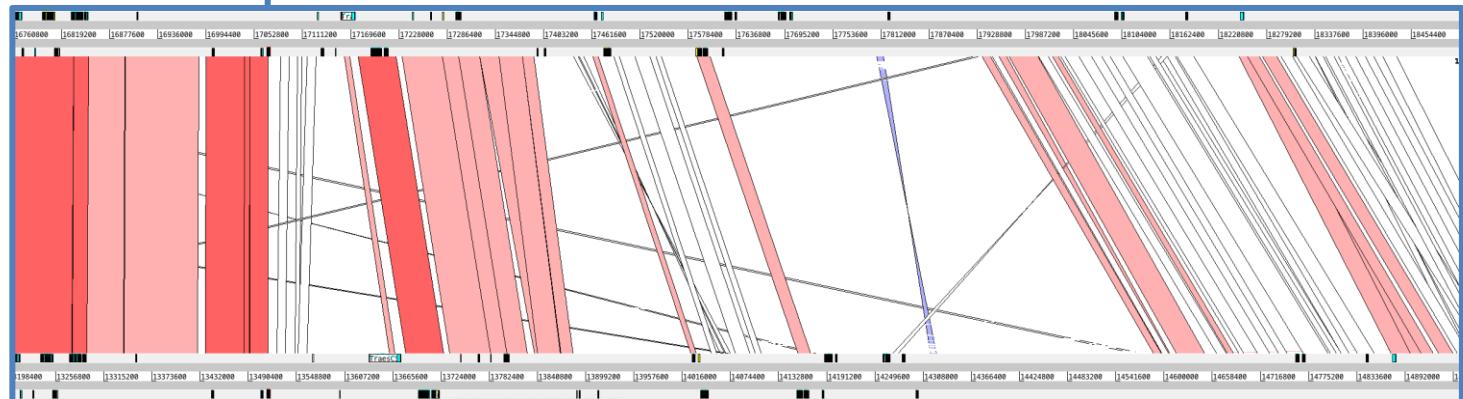
telomere



Chinese Spring – chr7D

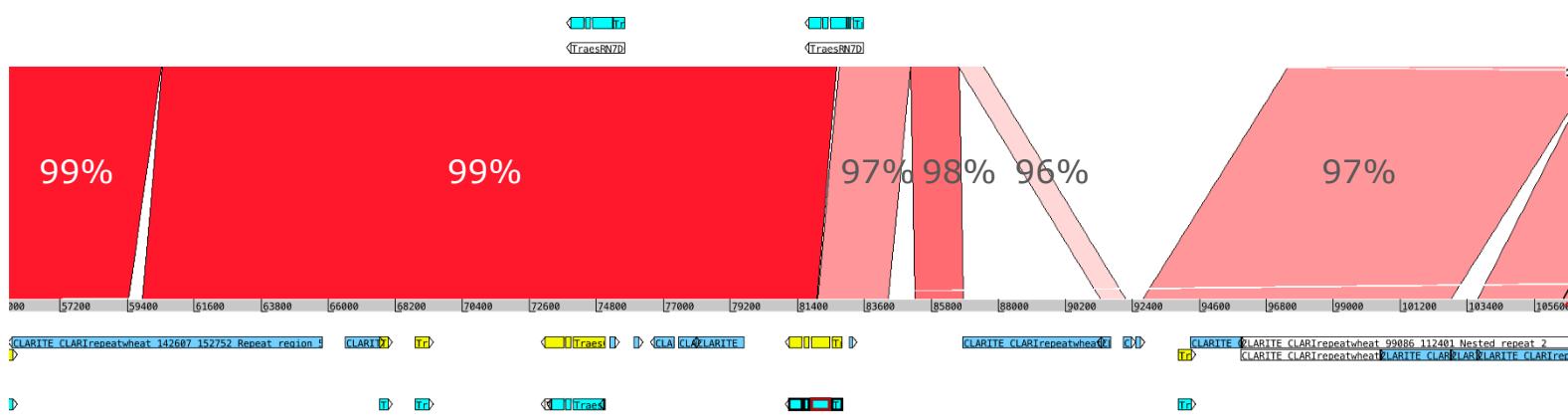
telomere

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>



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 Jamilloux³, Etienne Paux^{1,2}, Valérie Barbe⁴, Sophie Manganot⁴, Adriana Alberti⁴, Patrick Wincker^{4,5,6}, Hadi Quesneville³, 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



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

Thomas Wicker¹, Heidrun Gundlach^{2,3}, Manuel Spannagl², Cristobal Uauy³, Philippa Borrell¹, Ricardo H. Ramirez-González², Romain De Oliveira⁴, International Wheat Genome Sequencing Consortium⁵, Klaus F. X. Mayer^{2,6}, Etienne Paux⁶ and Frédéric Choulet^{1,*}

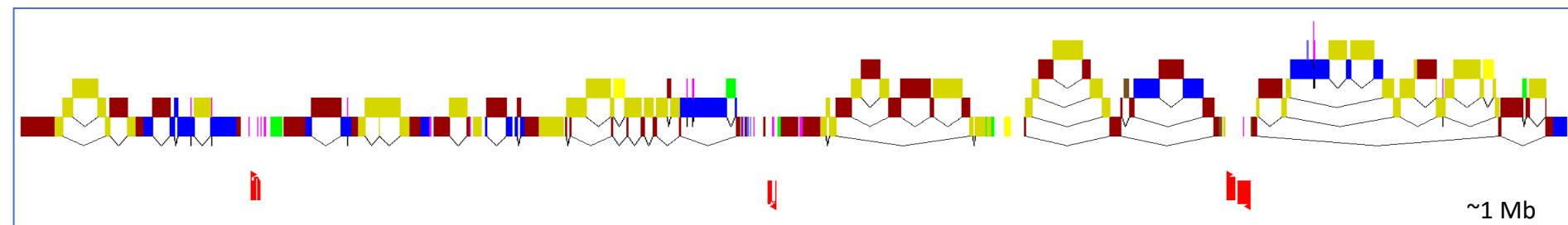
De Oliveira et al. 2020

Frontiers in Genetics | www.frontiersin.org

August 2020 | Volume 11 | Article 891

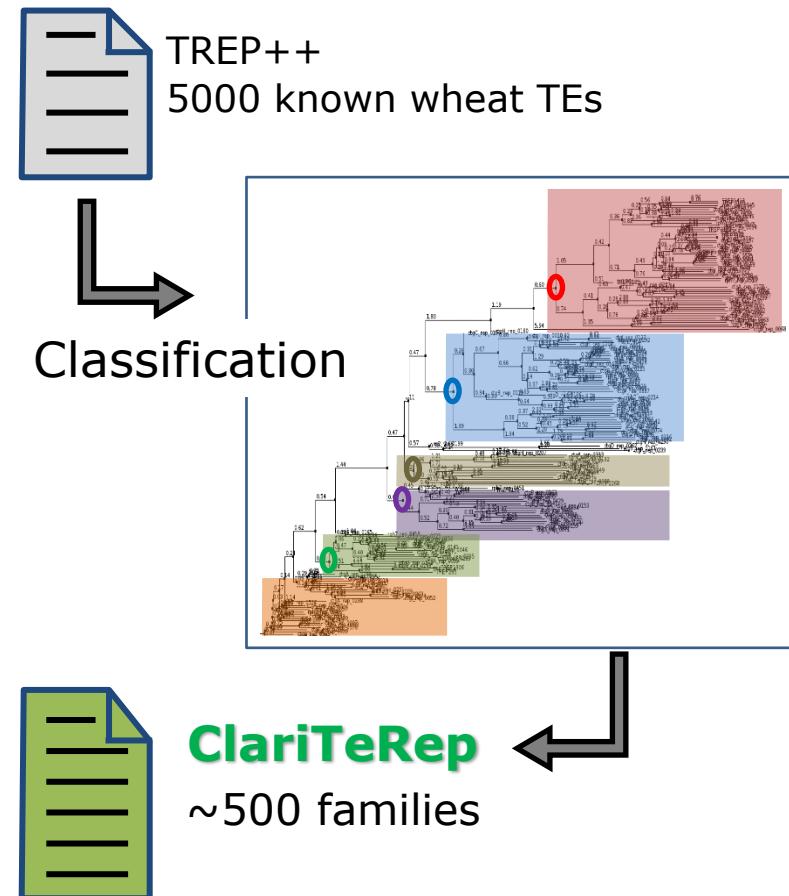
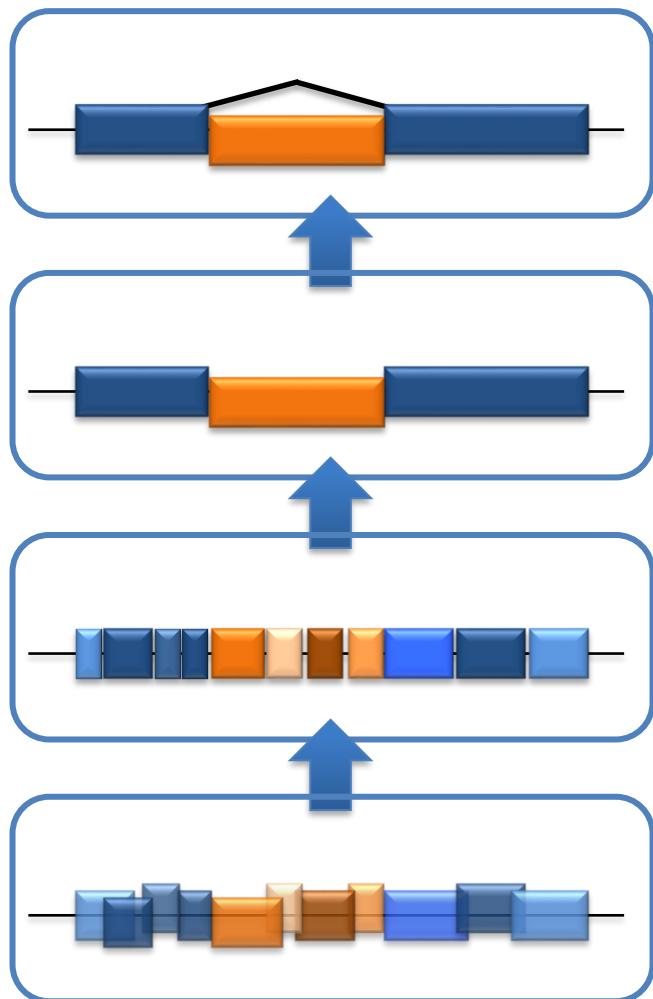
Structural Variations Affecting Genes and Transposable Elements of Chromosome 3B in Wheats

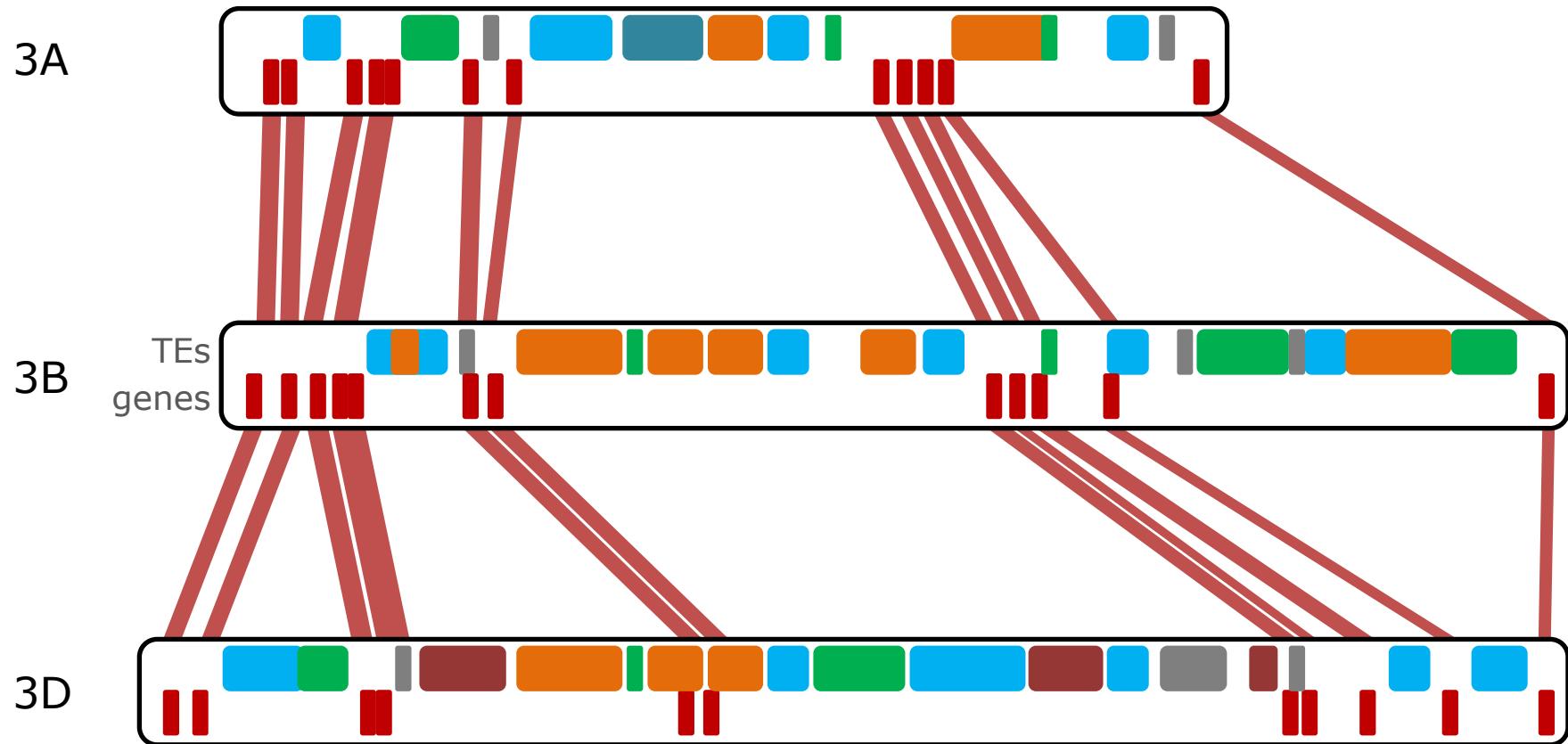
Romain De Oliveira¹, Hélène Rimbert¹, François Balfourier¹, Jonathan Kitt¹, Emeric Dynomial¹, Jan Vrána², Jaroslav Doležel², Federica Cattonaro³, Etienne Paux¹ and Frédéric Choulet^{1,*}



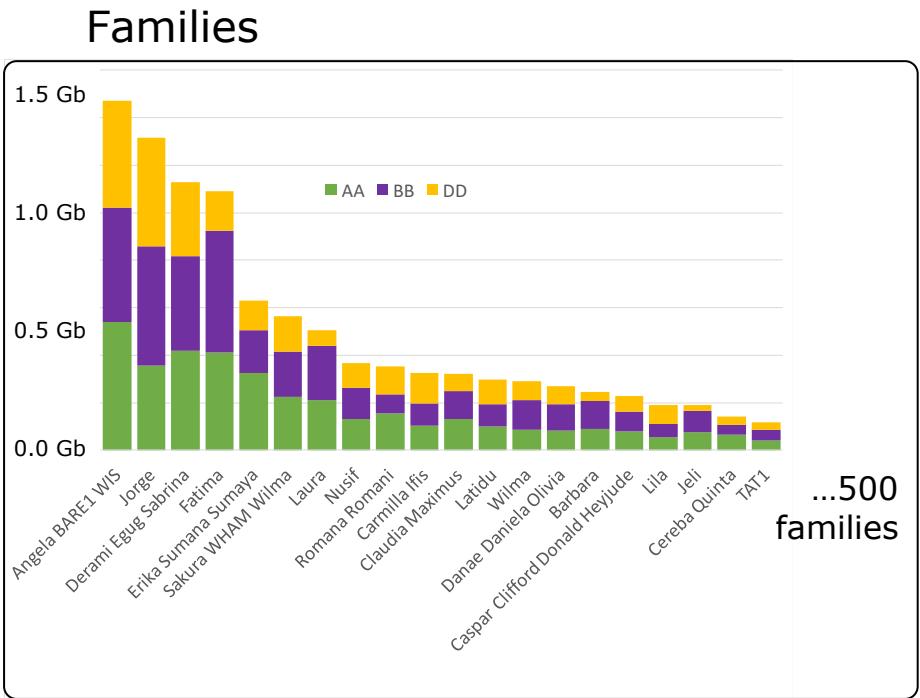
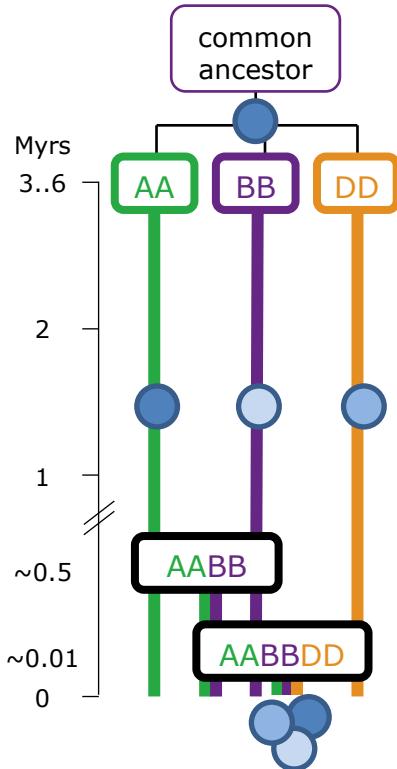
○ TE modeling with CLARITE and ClariTeRep

<https://github.com/jdaron/CLARI-TE>



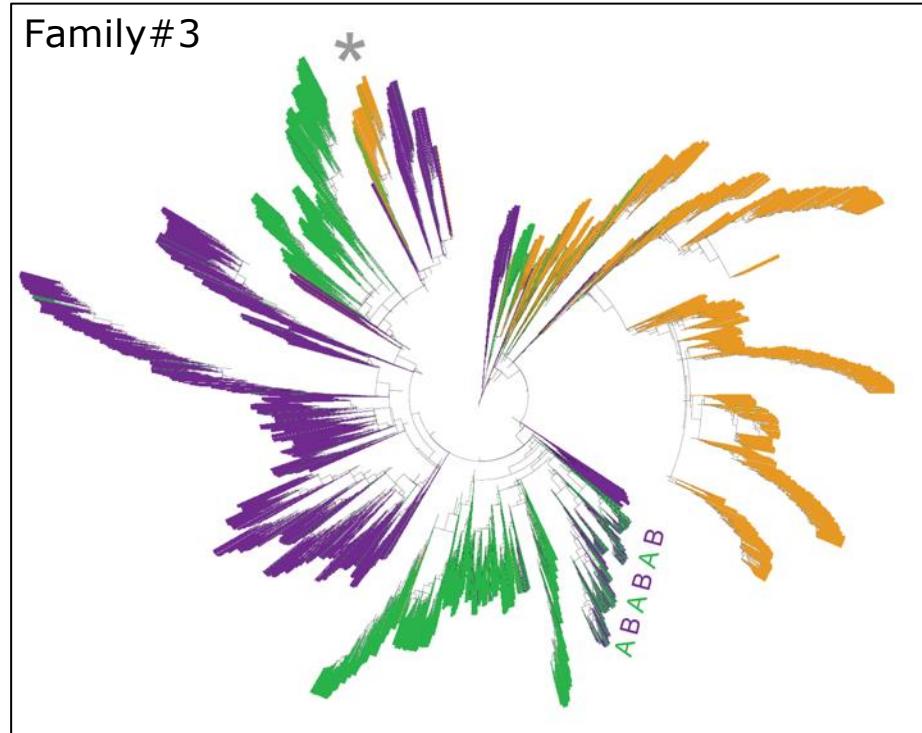
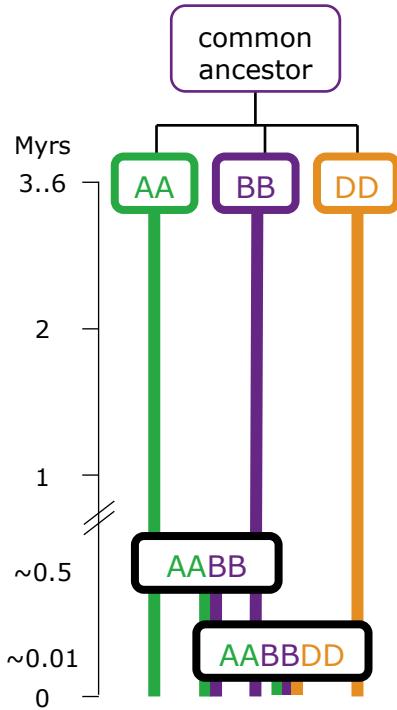


➤ Near-complete **TE turnover** since A-B-D divergence



➤ A-B-D balanced proportions

➤ 0 family specific to 1 subgenome

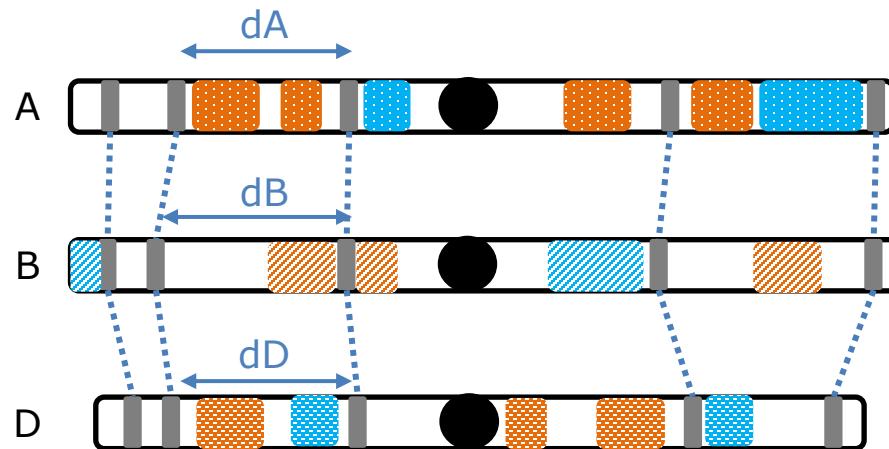


- Independent TE evolution in the diploids AA, BB, DD
- Some rare cases of TE amplif in the tetraploid AABB

A-B-D last
common ancestor



↓
~ 3 Myrs
TE turnover



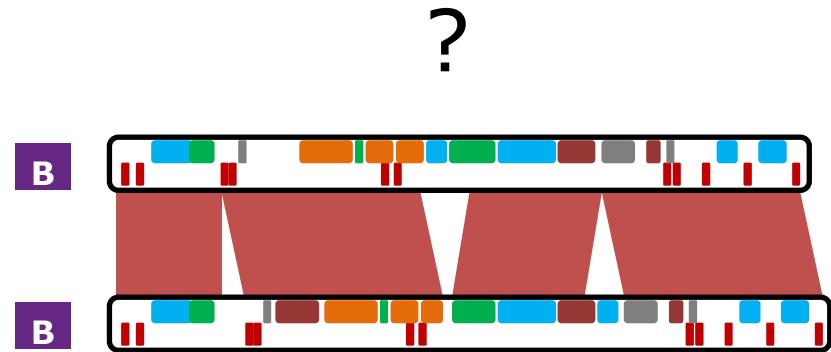
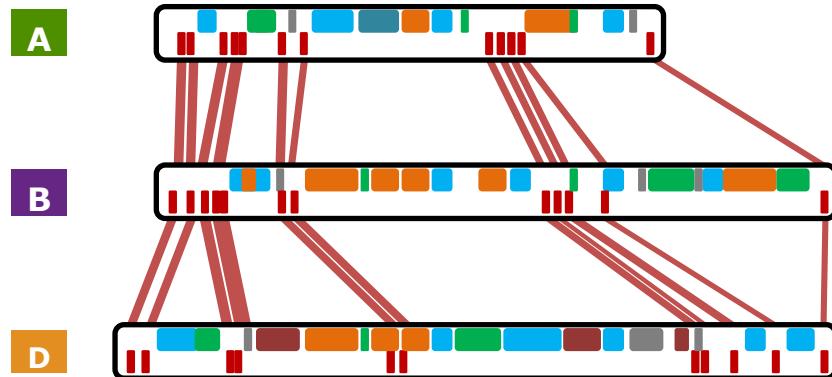
Most families were active since A-B-D divergence

➤ **Equilibrium**

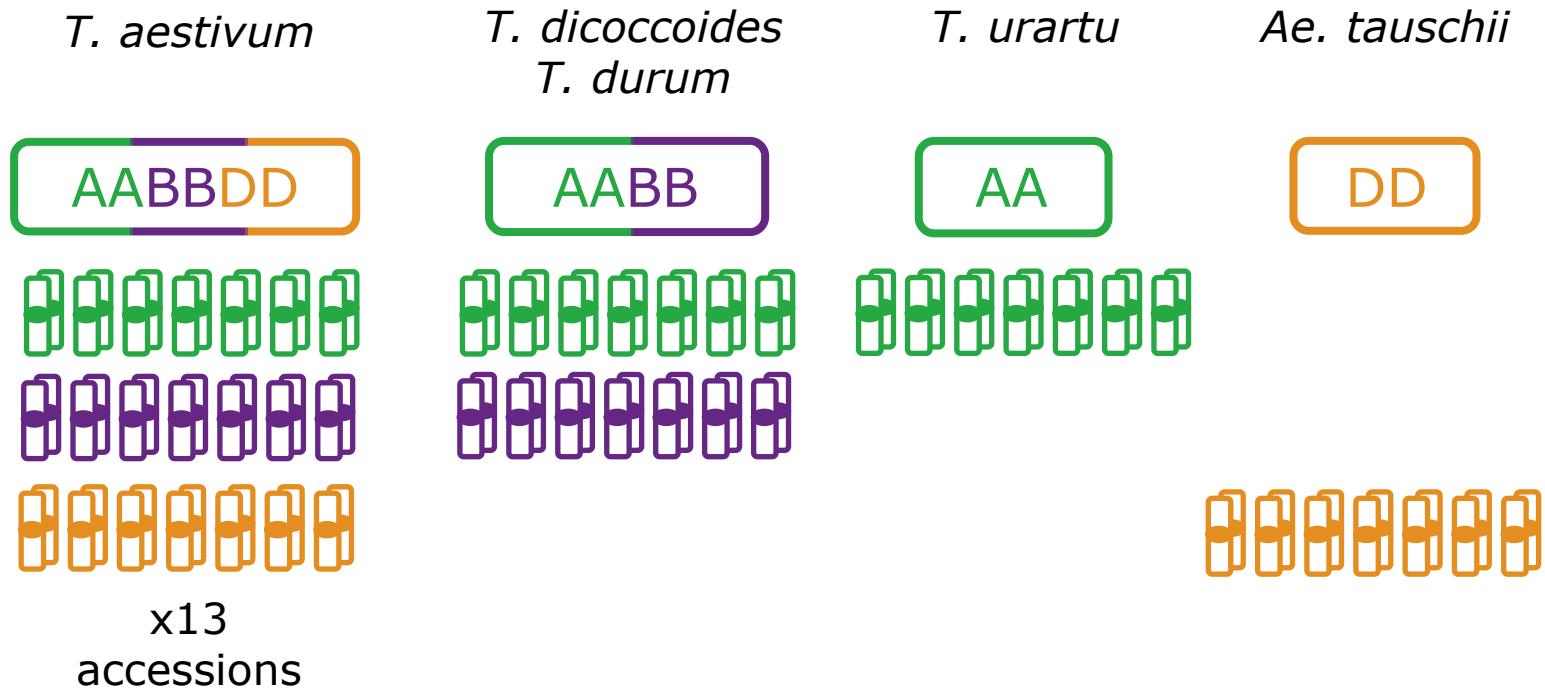
Hypothesis of a **structural role of TEs likely under selection pressure?**

□ Genomic variability affecting TEs in wheat **using assembled genomes**

PhD Nathan Papon (2019-2022)



- Material = multiple RefSeq now available



IWGSC 2018

Walkowiak et al. 2020

Guo et al. 2020

Aury et al. 2022

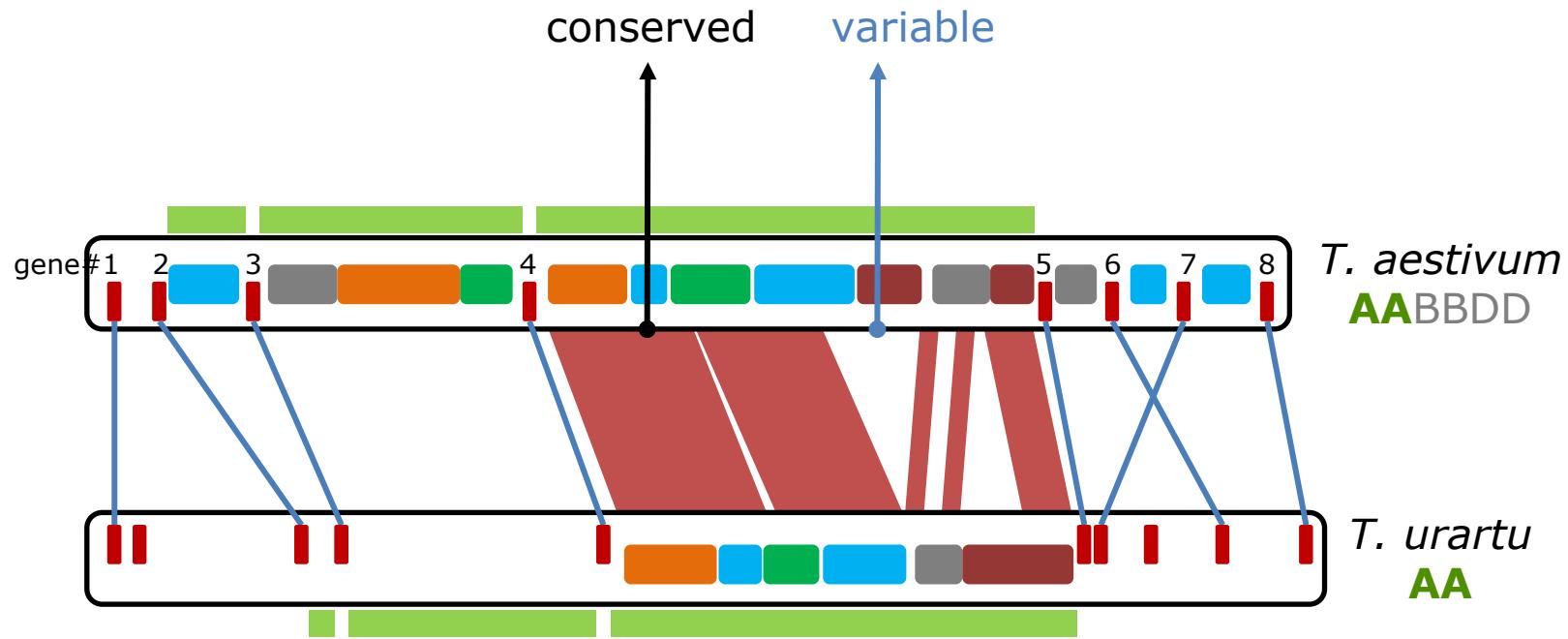
Avni et al. 2017

Maccaferri et al., 2019

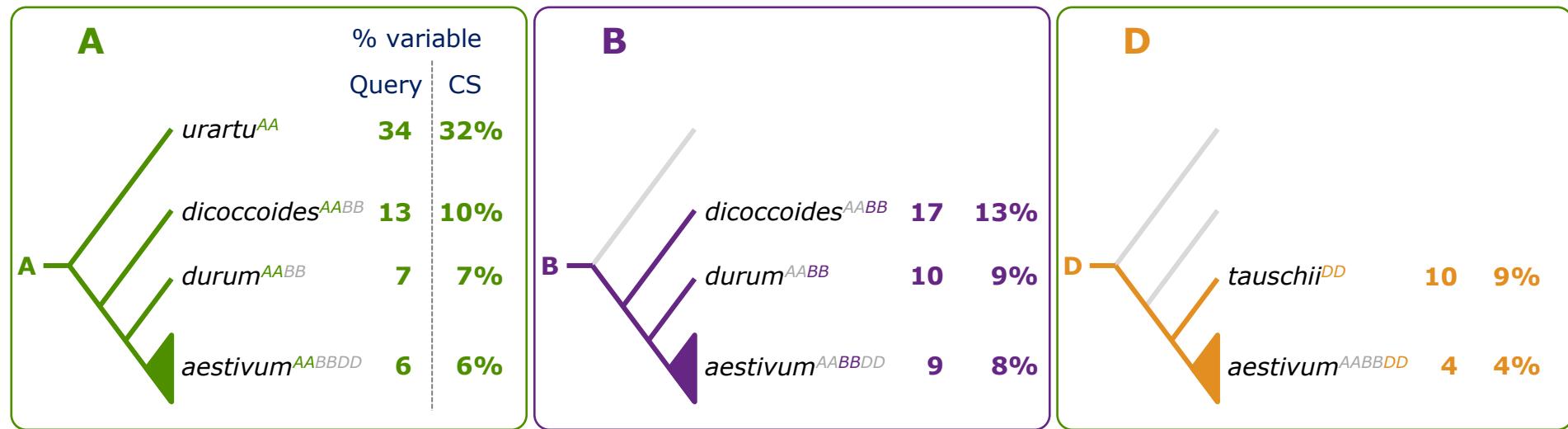
Ling et al. 2018

Luo et al. 2017

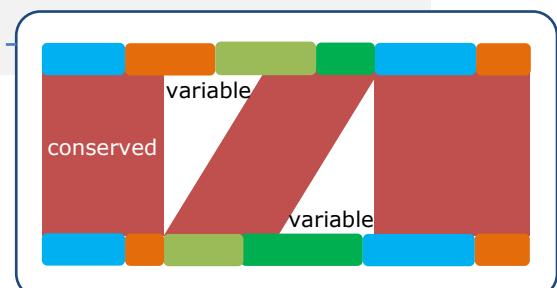
○ Method

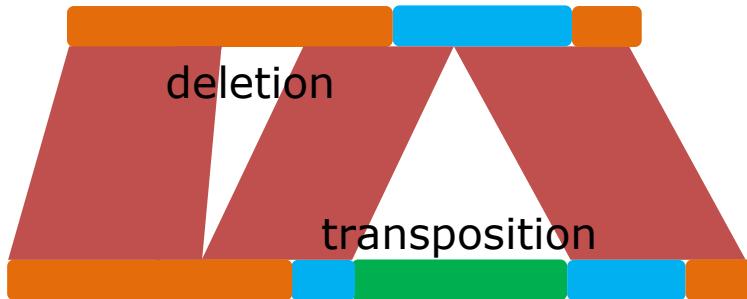


- Extent of variability affecting TEs



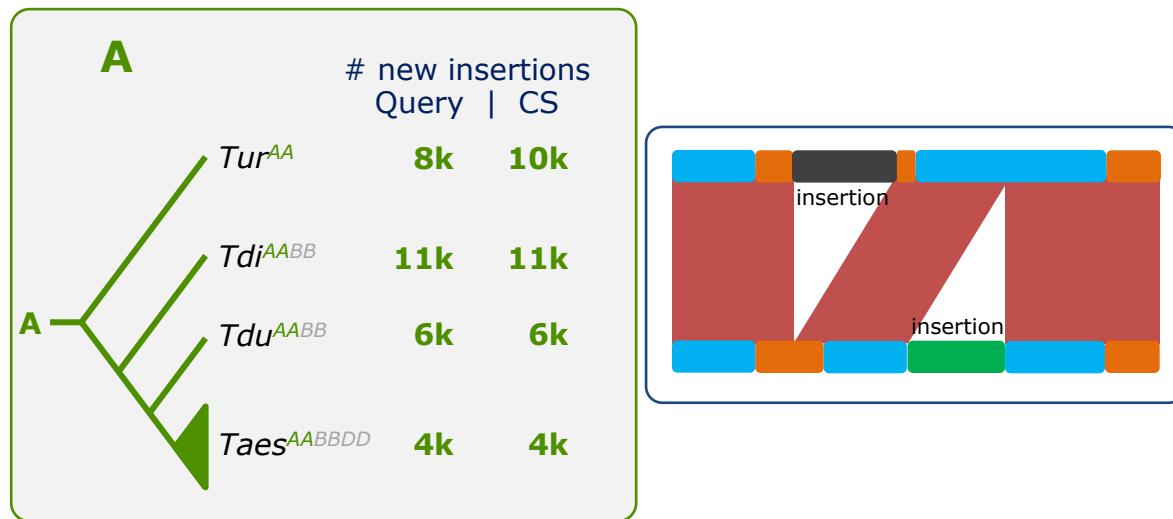
- 5% .. 15%
- B > A > D
- No genome expansion/contraction





Recent TE insertions

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



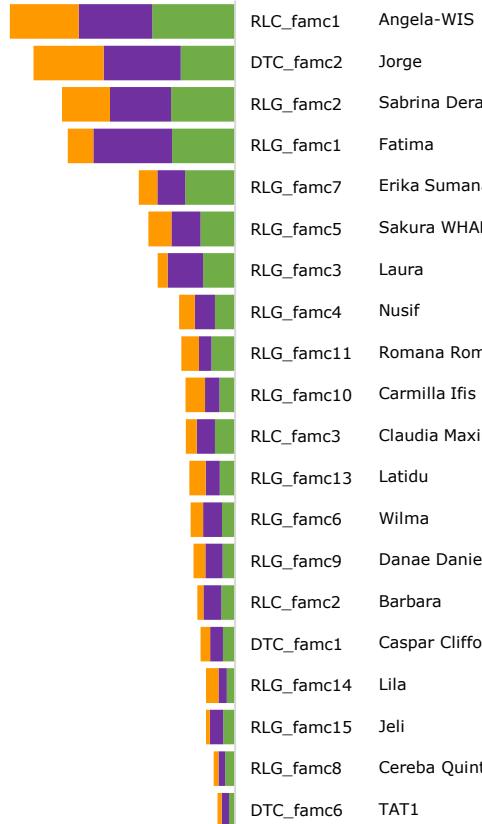
Impact of polyploidy?

- none

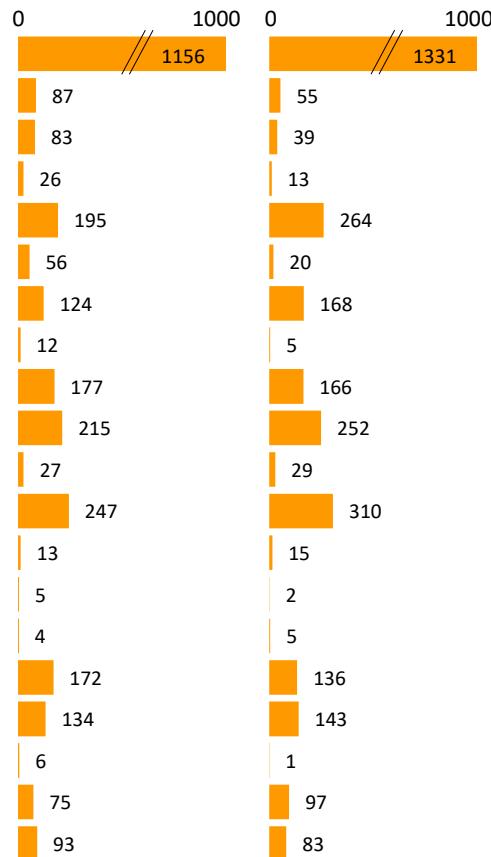
How many families were active recently?

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

20 most abundant families



#specific TE insertions



CS^{AABBDD}
vs
Ae. tauschii^{DD}

CS^{AABBDD}
vs
T. durum^{AABB}

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*



SRESRI/Région ARA



La Région
Auvergne-Rhône-Alpes