Genome Bioinformatics at GDEC

Frédéric Choulet

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

Genome Bioinformatics at GDEC

Frédéric Choulet


GDEC
GDEC
LA GLACIÈRE
CHANTERANNE
LA GAUTHIÈRE
Lycée Professionnel Roger Claustres
Musée d'Art Roger Quilliot
Basic-Fit Clermont-Ferrand
Usine Michelin de Cataroux
L'Aventure Michelin
LE VIEUX MONTFERRAND
Jardiland
METRO
Aulnat
Aéroport Clermont-Ferrand Auvergne
Parc Montjuzet
CPAM du Puy-de-Dôme
Polydome Clermont-Ferrand
McDonald's
1ER MAI
Maison des Sports de Clermont-Ferrand
CHU Clermont-Ferrand site Estaing
Leroy Merlin Clermont-Ferrand
Hôtel F1 Clermont Ferrand est
Casino#hyperFrais
Intermarché SUPER Clermont-Ferrand et...
LA GARE
Basilique Notre-Dame du Port
Av. Edouard Michelin
E.Leclerc CLERMONT FERRAND
Cathédrale Notre-Dame-de-l'Assomption
CTRE ANCIEN
Clermont-Ferrand
Rue Anatole France
LES RONZIÈRES
Bd Jacques Bingen
SABLON - TRUDAINE
LA PRADELLE
Centre Jaude
Musée Bargoin
SALINS
Jardin Lecoq
Stade Nautique Pierre-de-Coubertin
Bd Jean Jaurès
LE GRAND BEAULIEU
VALLIÈRES
SAINT-JACQUES OUEST
Basic-Fit Clermont-Ferrand rue de...
E.Leclerc CLERMONT FERRAND
Kizou Aventures Clermont Ferrand
Stade Philippe Marcombes
Université Clermont Auvergne - Campus des...
Stade des Cézeaux
Action
CHU Clermont-Fd Site Gabriel-Montpied
LE PONT D'AUBIÈRE
Ciné Dôme
Castorama
Jour de Fête
Beaumont
Auchan Aubière
Cultura
LE GOURGAT
Aubière
Google
Calques
A71
A711
A75
N89
D766
D769
D2099
D771
D777
D765
D772
Calques

GDEC

Research Teams

PaleoEvo

GeCO

DiGen

QualiGrain

Vista

MDC

Admin

Platforms

Bioinfo

Frédéric Choulet
Hélène Rimbert
Pauline Lasserre Z.
Vincent Pailler
Nathan Papon (PhD student)
Fabien Leloup
Jean-François Charmet

VegePole

CRB

Collection of 14,000 wheats

Gentyane

Sequencing Genotyping Platform

ValFon

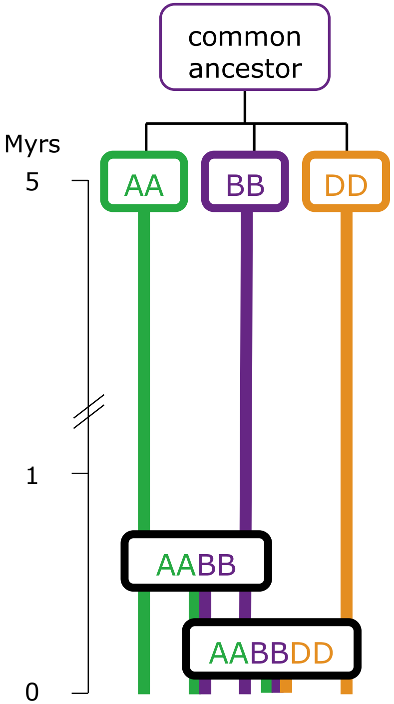
- Bioinfo for bread wheat and *Triticeae* (pan)genomics

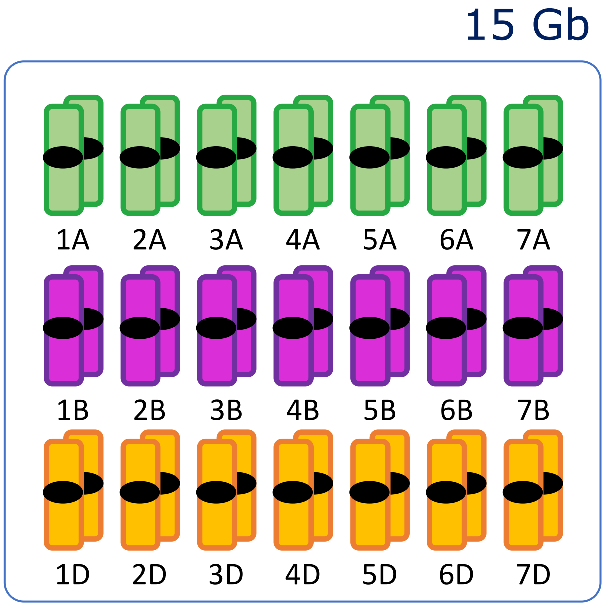
❖ Research area and questions

- 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?
  - …

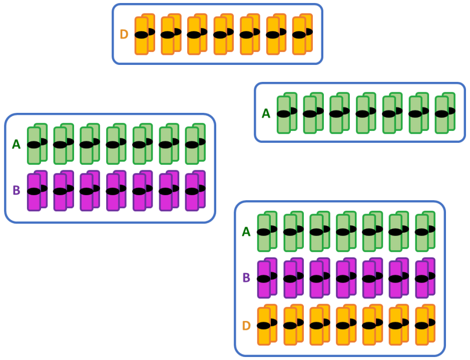
❖ Hexaploid genome





Triticum/Aegilops

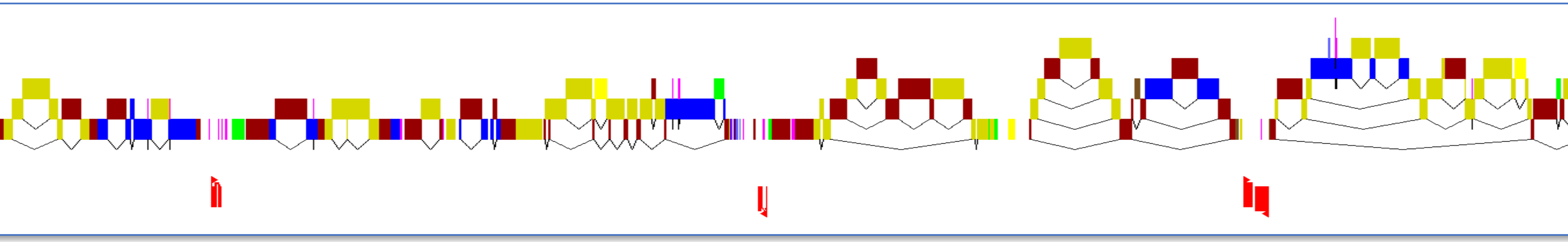
- Wild **diploids** (AA, BB, DD, SS, UU, NN, …)
- Wild & cultivated **tetraploids** (AABB, DDNN, …)
- Cultivated **hexaploids** (AABBDD)



CTGGACGCCGCAGGCATACCAGGGTGGTGTTGTGGATCTGATTTCTTCACCACACAAGCCCTCCGAAGACTTTGCGATGTCACAGGCAAAGGTTTAAGGCCTATTATGTTCTTCTCAGTAATCCTACATCACTTGGCATGAATCTAAA
TAAATTGATGATAACTCACTTTGCAGCTTGATTACAAACCCGCTGCTACATTGTCGGGTTTGGAAACAAATACCCCGAGGCATTGATCCCTTATGACGGTGTTAAGTATGGGTGCAAAGGAGGTACAGAAAGGAGAGTAAGAAGGCAA
ACCCTAATATCATCGTCAGAGCAATGGTTGCTGGTCCTGACAGGCATGATGGCTTCAAAGATGTCCGCAAGAACTACAACTGCACAGAGTCTTCTCTTGCACCAAATGCTGCTCCCTGCTCTGGTTGGCACTGCAAGCAAAACTAGTT
CAGTTGATAAGAACACAATCTTCTCAGCAATTCCCCCCATGCAAAGCAGAGGCCACAGACAGAAAGGAGACAACATCAGCAGCTAACATAGCTTCTAAATTCTGAACCTCTCGATAATTCCATGCATGATGAAATTGACGATCAGCAT
CAAACAAATAATCCGCCAATCATAGACTGGAACTTGTGTGCGATTTCTTAAACTAATTCTCGGATCCTTGTTATTCATATGCAAAGCTCATCTGCGAAGCCTTTTACTCAGATTGTACAGGCGCGACTTTCTCGTTGTAGTAAAAGAT
TTCTGTATCCTTCATACATACAGCTCGCTTCATCTTATTTCACTCATCTGTGTCTTCAATCCTAATACAGCACACCACTAATGGCCAATTCACTGTAGCCATCGTCCATGGTAGTTAAGCTCTGATATGCATGACGACAGAGCCTCCC
GTACCATCTTGTTGTTGCCCTCTACGTACAGGACGTCTTCGAACAGACTCTTTGCCTGCTCCAGCAGCCTTATCTTCTCCTTGTTGTTCCGAGACCGCAACTGGGACCGCTGCTGCAGCGTGCTGCCCCATCTGTAGAGAGCTGCAAC
GCCACATTGCTTACACATGAGCATCATACAAACAACCAGAGATGGAAAATGAAGTATCCAAGTTCACACTTGTAGGGAAATTCAGGCCAAAAAGCAGAGCGGTGTCGAACCTTCTGGAGCGTAAGTGTTGCTTTTGGATAACATGGCA
TCAAACTTGTCGATTGCAGCCAAATACACCCGGTCAGCATCTACTGCCGCTTCCTGCACCAGTTTTGATTTTCAGTAAATGTAACTAGAGTGCAGGTTTCAAGCCAAGTGCACAAATAGCAGAGCAAATTAACCTAAATACTTACCAG
CATGTTATTTTGTCGGGATGTGTTCAAGGATATGTAGTATTTCAAAATTTTAGGGTCAGAAACACTAGTTAAAAAATTACAAGAGGAAGTCTATGTTTATCAGCCTTAGTTCACCAATATCATGGGGCAACTATGTGACAATGCTAAT
ACCCTAAAAGGTGGTATTCAAAGACTATGACGTGATAAAGATGAAAACAGCTTGTT**ATG**TGAGAATTTGTATATTTGATACAAGTATGTTGATGCACATACCGGTCCGATGTCGGCAAGCAGTTGGCCACGAAAGATAAGCCGCCAGTC
CCCAATTGTACAGCGCCTTGACGTCGCCCGAATCAATTGACACAGCCGTCCTATAGCTTCTGCCAGCCTCAACAAGGAGGCTCTCACACTCCTCGCAAACGCCCACTAGCGCGGACGAAATGCTCTCTCTGTTAAGGATTCTTCTGTC
AAGCTTCCTGGACGGTGAGACACGCTCTCTTCCATTCAGATACGTCCCGCTGTTGGCCAGCAGCGTCCTCAGCTCGCGGCTAATCTTGAGCTTGAGCTCCCCATGGAGGAGGTAGGTGTTCCCTAGCTGACCGACTGCTACCAGGCTT
GTCGGCCTCAATGCCACGGCGGTGGACAGAAGCCTTGCTGACCTATACAGCAACGCGTCCGCGGCCTCTTCGTCGTCCCTGGCCATCATGCACTCCCTCGCCTTCCTCAAGATCTCCGCAGCCTCCTTGACGTTGCGGCTGAACTCTT
CAGCATCTGAACCCACAGCAGCACCACCACCCTCACCATCATCTGCAAGTGGCTGCTCATCGACACTGGCGGCAGCGCTGCCCACAAATTCATCAGCCGGCATCCCCCGTGCTCCACTCGCTCGGCTCTCACCGCCGTCGCGGCGCTC
GATCTTCAACCTGTACGACCTGTCCCGGATCTCGAGCGTCCTCTCCAGCAAGGAGGACTCGGCGGATTCAGCGGCGGCGAACCTGCCGCCTTCTTCAAAGCGACCGTCGCTCCTCGCCCTGCCGCCTTCTTCAAAGGGGCCATCATTT
TGCACCCTGCTGCCGCGGCCGTGCCTGTGCTTGGACACCAGCCTCTCCAGCCGGTCGTCGTCTTCCTGCCCACTCTCAGAACTGAACCCTGCATCCTCAAACGCATTGGCAGCATCATCCCTGCCATCAAATGGCAGCATGTCGAGCG
GCCGGGCGCTTCCGACGGCCGCTGAATCAACTGGCTTGTTTTCCCCATCCCCGTGCTGAGCTACAGCAGAGTCTGAATTACTCTTGGCATCGGCCG**TAG**AAGCAGCATCAGGGGCGGCGACGGCCTTGGCGGGCTTCTTCCGCGGCGA
GGCGAAGCTGAGCTCCCGCAGGGCTACCCGAGCCCAGCCGCCGATGTAGCCGGCGACCTCCCTGGCCACGTCCTGCAGCGCCCCGTCGTCCGGAGCGGCGCCGCCGGCTTTGGTGCTGCCCTGCCCCTGGTCGACCACCGCGGCCGCG
CACTCGAGGAGCGCATCGCGGGCGCGGTCGAGGGGCTCTAGGATGGGGCGGCCCCTATCGTCGTCGTCGGCCAGGAGGCGGTCCTTGAGGGAGAGGAGGCGGGCGTCGAGATCCGCGAGGAGGGCGGCGGGGCGCTGGCCGGAGGAGG
GGGCGGGGGCGGGGGCGAGGTGCCTGGCGGTGGCGAACCCGGCGGCGAAGACGGCGGCGAGGAGCGGGGCGGGCGGGAGGCGGGAGAGGGAGAGCGCGGCGACCGGGAGGAGGAGGAGGAGCGGGGCGGCGGCGAGGGATTGCTGCGT
GACGCGGAGGAAGGGGAGGCCCAGCGGGTCGTCGGCGGCGGGGAGGTCGGCCCAGCCGTTGTAGGACGCGGCGGATGAGCAGCAGCAACGGGTGGATGTGGATGTGGATGTGCGGGTGGAGGCTCGATGCGGGCGGGATCGGACGGGG
AGGAGGGGCCGGTGGCGGAGGCGGGAAGGTGTCGGCGGCCTGGAGAGGAGGGGGAGGAAGAGGGGCTGCGAGGAGGGAGGAGAAGGGAGGGCAGAAACGGCGGAGGATGCCGCCATTATCCCACCCCACCGCTGCCTCCTCCGCCTGT
CTACCTTTTTTCAGTGCGAAATATCCCACCATGTCACAATTCTGCGTTACACTGATTCGCCGTCTTTCTTTCTTATCAAATCTATTATGAAAACTCATCAAAAGTACAAAAAAAATCCCAAATATAATAATAAAAAAACACATCAAGA
TTTCAAGACCACCCAACAAACATTTGACGTCGTGACGTGCCGCCGCCGTCACCCGTCTACTGAATACGGCTTGACCTTACCGATGATGACCTTCAAGTCCTCGCGCGCGTGCCTCCAAGGAGCAGCGCCCTGGAGCAGCAGTCGGCGT
CGTCGGCGTTTAACCCTTGCATAGATCTTAAGCACCTGACCCCAAATCTCACACGTGACAAGGAACCCTAAGCTCGCCGCTCCAAACACACGCCAATCTATGTCGGAGCTCCGTCAACTTCGTCCACCTGGACAAGCTCGACGAAGAG
CAAAGCCTGGAAAGACAAACTTGAAGAAGGAGAGTCTCCATCCCCCCGGACTGGAAAACCTAACCTTAAACTATTAATGCAAGTGAAGGCACTGAGAGTCTCCCTCCCTGCTACTAGCCGCCGAAGTTTATGCCCTGAGAGTCTCCAA
ATAATTCTTGGATTCACCCCTTTACAACAGTAAAATCTCTGCTACTCCATTGGATTACTCAAAAATCTATTATAAAAAATAGTAAAATCAAAAGAAATGATGGAAACTATTTAGAAAAACTGGATTTTCTTTCGTAACATTCAAAATC
AAAGCTACAATTCTTAAAAACAATGTTTAATGAATAGTAGGGTCCCAAATTGATTTTGTTCACTAAGAGTATCACGGGTGTCATTACTTACAAACTTCAGCAATTTTGGTACTTTAAAATTACAGAATTTTTTATTCTTTCTATTTTT
TAAGTAGTAGACATTTTAAGAATTTGATCGTTAGGACCGTTTAGGGGAATTTTTTTTAAGGTAGATCCAACTACCCAAAAGAATGGACTTACACCAAAAAGATAATTCGGTGGCCATATGAATTTTAGGTGACTTACATAAACAGAAA
TCAATTTATGTGAAACTGCTGCGTTTCCCATCCGTGCAAATTTGAGCAGGGAAGTTTGACAAAAATAAAATTTAAAGGGGTTAAAGTATGTTCTACAATAATTTTAACACAAAAAAAGAAAGGATGCAACAATAATACAGTAGCATTC
AAGCTACAGAAATGCCAACCACATCATCACAGTATCTAACTCCACTTCCATAAACAGAAGACAGCACTGAATTTGAGATGAAAAACCAAATGATAGTAATATATGCAACATCAATCAACAGGTCTTTCCAAAAAAATCCAACACTTGT
TATAAATATACAAACACATACAGCAGCAGCACAGCCTTACCTTATTCTTACCAGTCCCCTCTTTCTCTCTACAAAATGCAATGACAATATTAACGAATAAAGTTTGCTGGTCCCATAAGTGACAACCCTCCTGCTCATAATAACAAAC
TCAAATTAATAATGGTACACTGCGTGTGTCTCCGCCATGATTCCCCGTGGTGAACTCAAGACGATCTCACCACCTCGCATCGTATGTGGCCTTCACGCTCCTTCCCCATCTCTCCTGGAACAATTCTACTGACACCCAAATATAACCC
TATTGTGAAGGATTAGCAGCTGACCGAAGCCACCCTCGTTCAGGCAGCGGCGACGGGCTCAGCTGGTGGCACAGGGGCTGCTGCTGCTGCGGGGGCAGCGGCGGCGGGTGCAGCGTTCTCCTTCTCAGGGACAGTGCCGTTGGCA
GGGGCATCCTTGGCAGCAGGTGTCGCTGCCGCCTTAGCGCCGTCCTTGCCGGCGGCGGCCGGAGCTCCGTCCTTCCCGGCAGGTGCCGCGGCGGGTGGTGGAGGTGGCTTCATGTGTGGCGGCGGTGTGTGCTTCAGGTTCGACAGGA
TCTGGCGCATCCGTGCCATGTCTGTGGCCTTCCCGGGCTTGGGGCCCTGGACCACGATGATCCCAGGCTTCTTGTCCTTGTCCTTCTTGCCGGTTCTCTTCTCGATGTGCGCAATCTCCTGTGGGATGAGCTCGGATATCGCCTTCCA
GTACTGCTTGTCTGCTCCCTTGTGGAACTTCTCCTGGTTGGAGAGGAAAAGCTGATAGCAAGCAGTAACACAGGGAGAGAACAAAAGGAGATGTCAGGTAACCATATGTTGTTCGAAAAGTGCAGAGAGAATACACAGGATCATGATG
TTAAGAAATACTCTACCTTCTCCCTTTCCCGGTTGTGATCCTTGCTTGTCTCAAGGTTCAGCTTGCGCTTCTCGACGAAACCCTTCTTGAACTCTTCAGCTTCAAGGATGATTTGGCTCCTCAGCTCCTTCTCGTGCTTCTCCTTTTC
CTCCAGGAGGAGTGCATTTTGACTGTTCACAGCAGGTGTTAGCATTTACTTTTGCACAATACAGATGAAATCAAGTAGTAAAAATTGCAGTACTTATTCTTTAGCAGATATGGCCTCTGCTCAAGATATTTAATGCAACACTAGTGTT
GTAGTTGATGATATCGAAAACAGGGCCTACAACGATTTATATAAAGACATTCTCTGGTCCGGAAGTCAGATCTGGAGTCATATGTTACATCTTTATCTGAAAGTTTATTTCGTACTGGCCACAGCAAACTGCATACTTTGGTGCAGGG
GTTTAAGTAGACAAATAGTGCACTGGAGGAAACATAACAGAAAGGATAATAGAAGCAAACACTTCCTATAGGGGAAAATAACACCAACATAATCAAGAACACTACTTTAGTTTTTGGAAAGCTCTCTTTTTGGACACCGCACTTTGAC
TACCTACATATATGTCAATATTTATATTGGATAACTGGAAGAGGAAGTCAAGTGTAGATTTGTCTCAACAAATAGCCAGTTTGAGAAACCAAAAAAATAAACTTGAATCATACAAAAATACAGGCTTATGTGTATTTGCCACTATGAG
CTAGAAATAAGAACAAGGGAAGTCAATTTCATTTTTCAATTCCCTTTAGATAATAAAACTAAAGAGGCAGCTACACACAGAGAGTCCTATAGATTCAAACTTGCAGTTTCATCATGCAATTGGACCTTCTTGATTTGATATCCTGAAT
TACAAAATCCCAGGTTTAAATAAAATTGTGCACCTAATGTGCCACAAGATCAGCTAGCCTGAGAAAATGTTCCAGCAAGATGAATCATTTTTATCATACTACTACAAGGTCATGATGGAAAAAGTTATGGCAGTGACACAAACATATT
CCTAGTGCAAAAACATGACTTCGGTGACCAATTTGTATGGCAATTGCTCCTAAGCAGCTAAATTTCTACTGCAAAAGCCGCAAAGGCAGTAGATCACATCATCATATGGCAGATTCCAAGCAATCATTTGCTTTTATTCAGCTAAAAT
CCTAGTGTACTGAGGGTGCGAAGGCACTGGGTCACACCATCATACAGCAAGTGCACTGAATTCCAAGCAATCATTTATTCTTGTTCAGCTTAAATCCTAGCGTAGTACTGAAGCAGCGAAGACCGTGGATCACATCAACATTTTCATT
TTCCTGTTGGCAAAACCTAACACTGCAGTCACACCCAGAGAGTCCCAGGCAGTAAAAATTGGAATTTGATCATGTGATTCCGGCTGCGCTTTGTTGACCTGACACCACCCACTACAAAATCCCAGATTTAAAATACCTACTTAGCGTG
TCTCAAGCTTAGCTAACCGTGAGAAGAAGTTGCAGCAAGCTGAGCCCTTTTTACCTCATTAACAACAACTCATGGTTCAAACTGCTGCTTCGATCACTTCCAAGCTACTAACCACGAGTACGGTGACCGGTTTGCACGTCAATTGCTC
TAATTCAGCTAAAATTCTAGTGTGCTGAAGCTGGAGTCGCTGGATAACAGCATCACATGGCACGGGCACCGGATTTCGAGCAGGTGTTGGCCTGAGAGACCGAATCTGCACCACAACGATGGAATCTCGCGGATACATCAGCTCAGCG
CCCCAGATAATGGCAATGGCAAGCGGGAGCCCAGATCGCGACCCAGCGCTGCAGATCTACGGCCCCCACGAACCCAACCCGCCACGATCGCGCGGCTAGATCCAGCAGGAGCAACCACCGAGCCGGGCGAGATCGATCAGATCTGGGC
GAGAGGAGGAAGGAGACGGAGCGCACGTACCGGCGCCACTCGCGGAGGAGGAAGCCCTCGTCGGCCCCCATCTGGGCGGGGTCGGGGAGGACGGGGCCGGGGTCGGTGAAGAGGCCGTCGTCGGCGTCGACCTCGTGGTGCATGCCGT
TGCCGTTGGAGTCGGGCATGCCGAAGGGGTGCGCGGCGGCGGAGTCGTGGCGGAAGGCGTAGGGGTCGGGGGAGGCCGGGAAGTCGTCGGCGCCCCCGGAGACGTGGACGTCGGCCGCGGCGGGGGTGGGGGAGTCGGCGAAGCTGAG
GAAGGGGTCGTCGTGGAAGGACGCGGCGGCCGGGGAGGAGGTCGGCGAGTCGAAGGCGGTCGCCATGGGCGATGGGGTCGGGATCTGGGCTGGGGGAGGGGGGAGGGTGGTGGTGATCTGTGATGTCGCGTGGGGGGGAGGGGAGGTG
GAATGGTGGTGGCGATGCGGGTGGTATATGACTCGTGGTGGGGGGACGGGAAGGTCGTTGCCGCCCCAAATTACTGGACTGCCACTGCCCGCCCACCTCCCTGCCAATGGGCCAACCATCATGTTGTTAGCAGCCCATGGTAACCCTG
GTTTATTCGGTTCATTCGGTTTACATGGTTCGGTTTATACGGTTGTAGCTCATCAACCTCGTGGTTCCAACATCCCCATCATCTTGTAGCACCAGCCACATCCCCGATGGCAACAACCTCGTGGTGACCATCTTGCAGCTCGCAGGCC

❖ Genome composition

- ○ Transposons (85%)
- ○ Genes (2%)

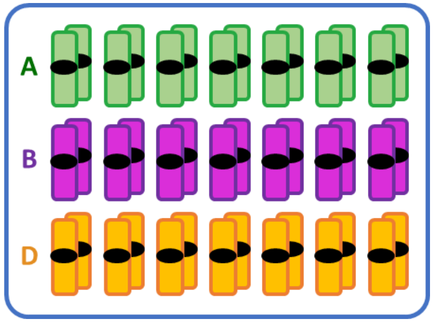


1st genome seq assembled:

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

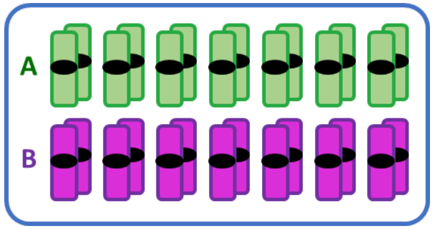
- ○ 4 million TEs
- ○ ~107,000 genes

❖ Pangenomics

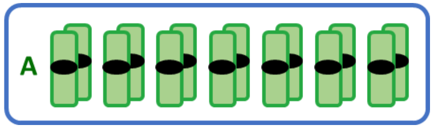


- *T. aestivum*

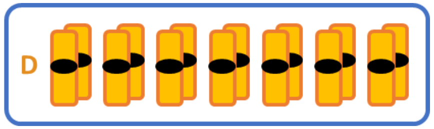
▪	**Chinese Spring** (Illumina)	*IWGSC 2018*
▪	10 other varieties (Illumina)	*Walkowiak et al. 2020*
▪	Zang1817 (Illumina)	*Guo et al. 2020*
▪	Fielder (Pacbio)	*Sato et al. 2021*
▪	Kariega (Pacbio)	*Athiyannan et al. 2022*
▪	Renan (ONT)	*Aury et al. 2022*
▪	Attraktion (Pacbio)	*Kale et al. 2022*



- *T. durum* AABB
- *T. dicoccoides* AABB

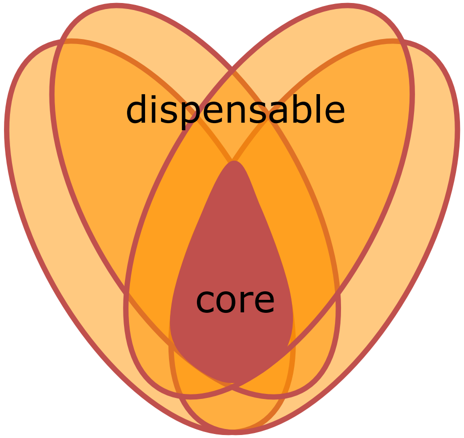


- *T. urartu* AA



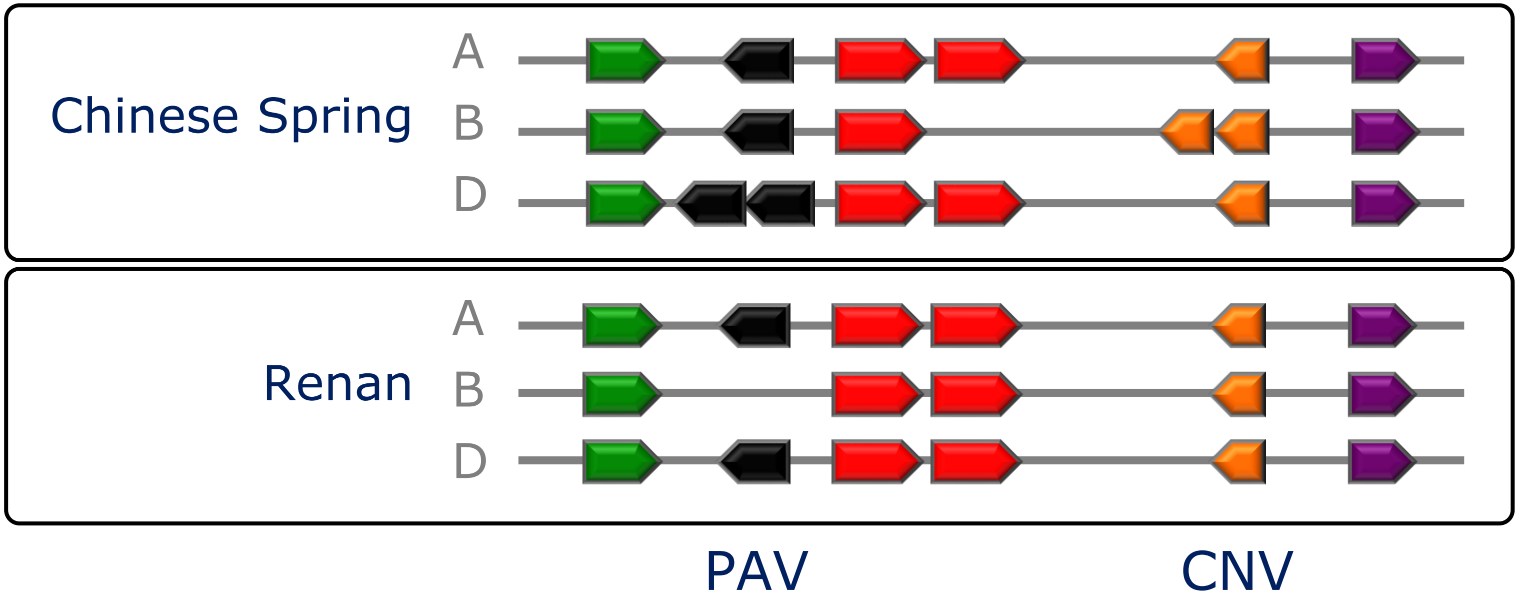
- *Ae. tauschii* DD
- *Ae. searsii* DD
- *Ae. bicornis* DD
- *Ae. longissima* DD
- *Ae. sharonensis* DD

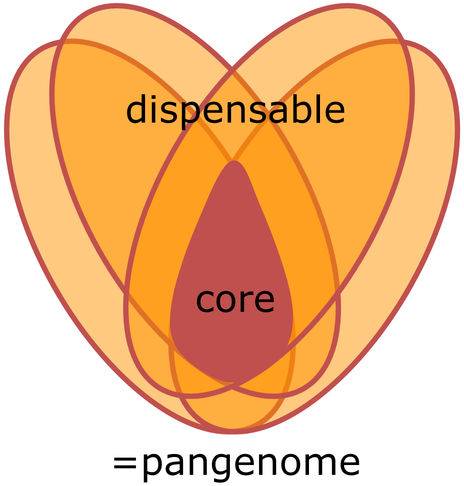
- *Ae. speltoides* SS



=pangenome

❖ Pangenomics





- ➢ Chromosome-scale assemblies required
- ➢ Quality of annotation matters

❖ Renan

OXFORD (GIGA)^n SCIENCE

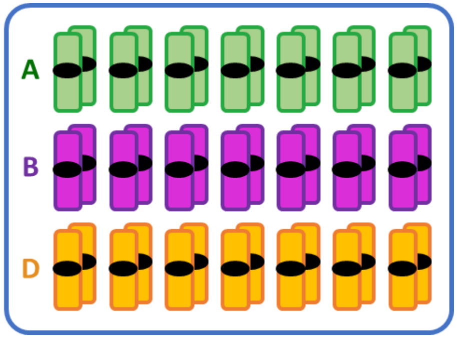
GigaScience, 2022, 11, 1–18

DOI: 10.1093/gigascience/giac034

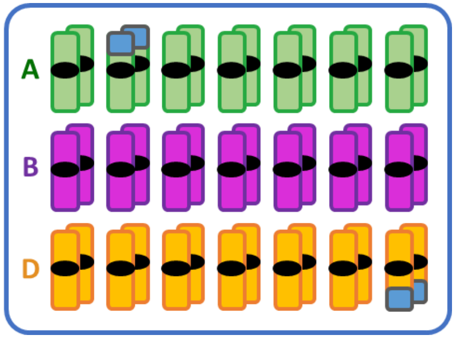
RESEARCH

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,*]

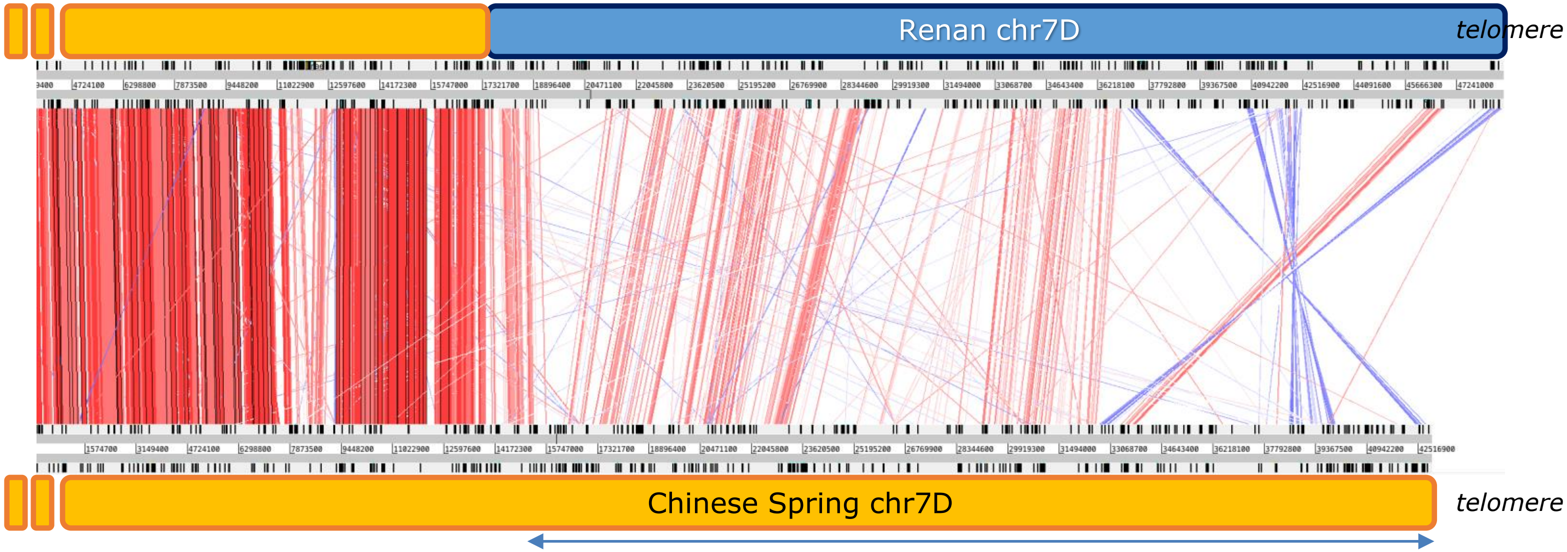


Chinese Spring



Renan

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



30 Mb, ~500 genes

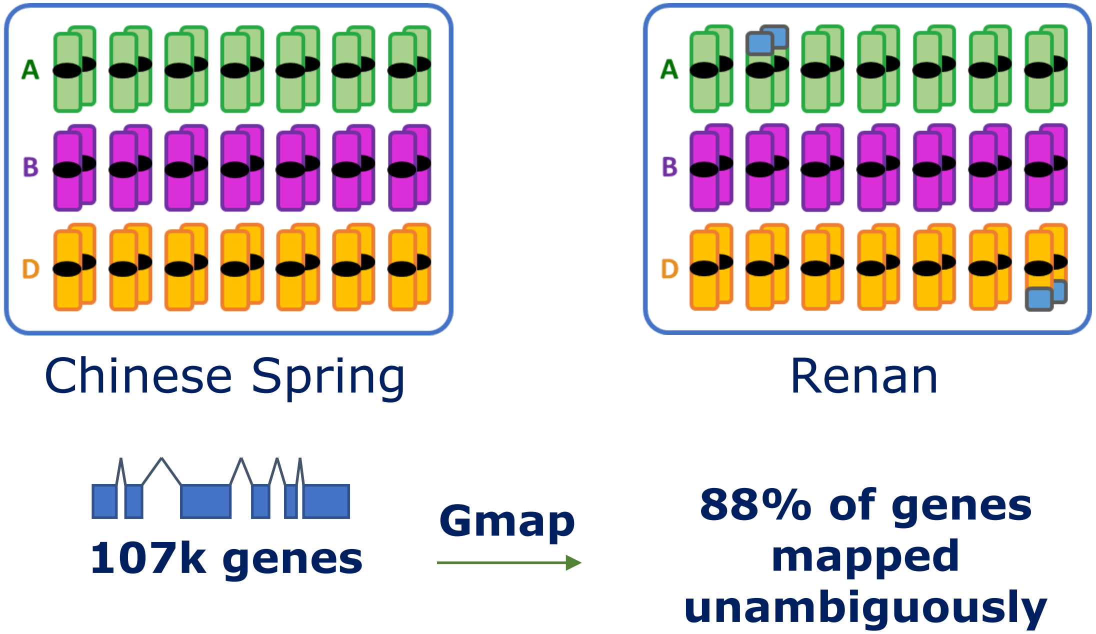
❖ Genes

❖ 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. Rimbert)*



Chinese Spring

Renan

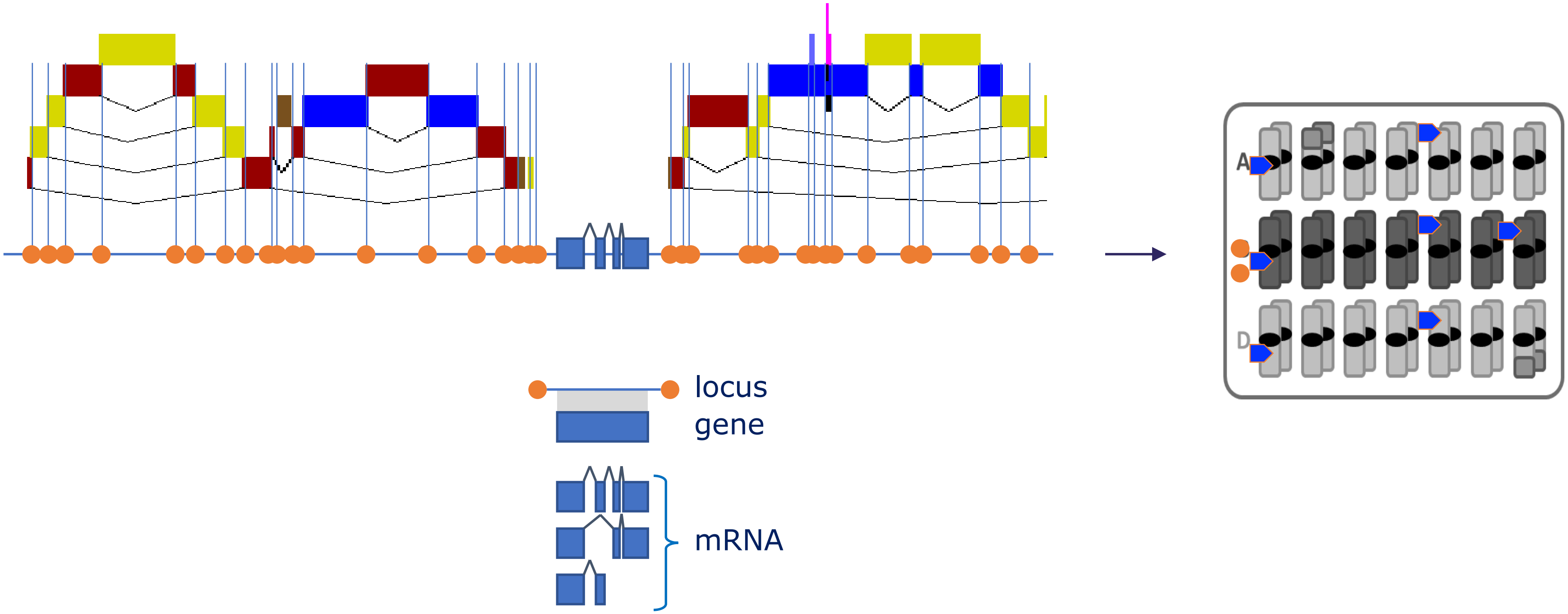
107k genes

Gmap

88% of genes mapped unambiguously

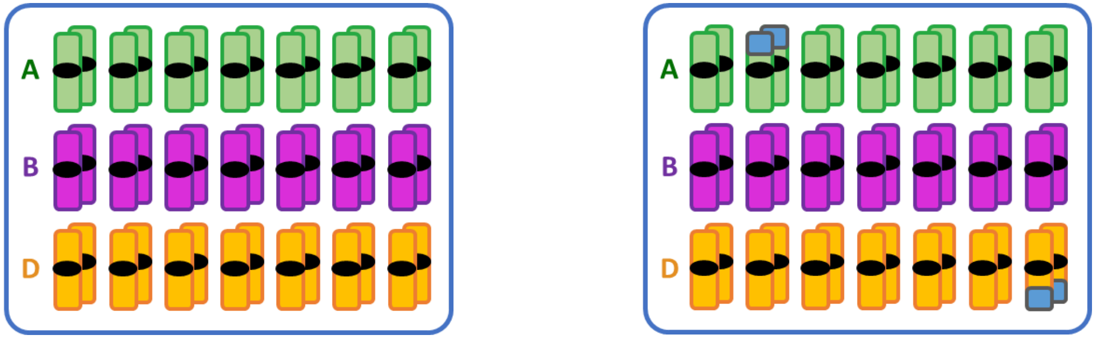
❖ Gene annotation

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



❖ 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. Rimbert)*



Chinese Spring

Renan

107k genes

Gmap →

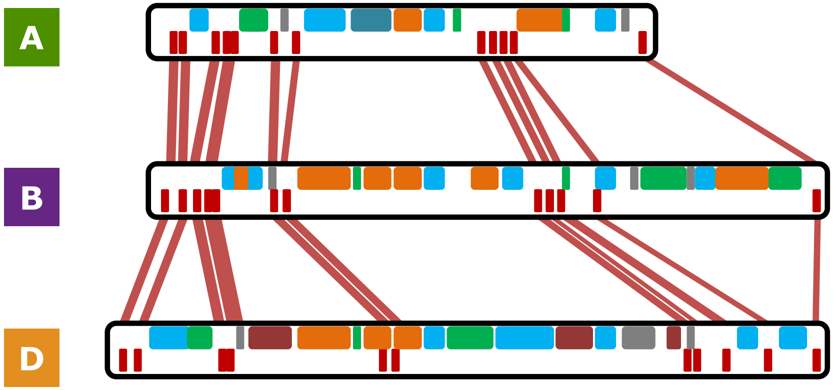
88% of genes mapped unambiguously

MAGATT →

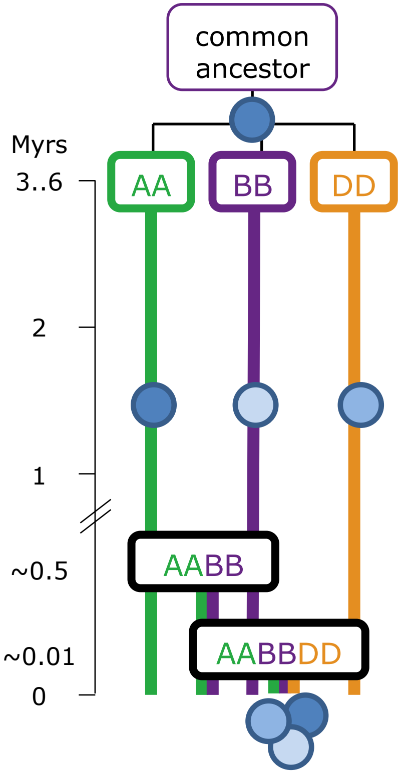
98% of genes mapped unambiguously

❖ Transposons

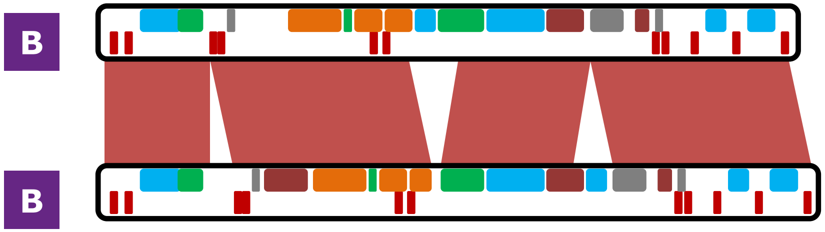
❖ Transposon dynamics

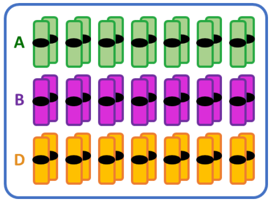


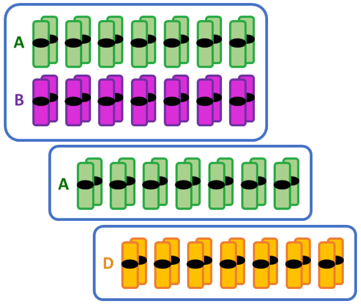
➢ **Equilibrium** model of evolution



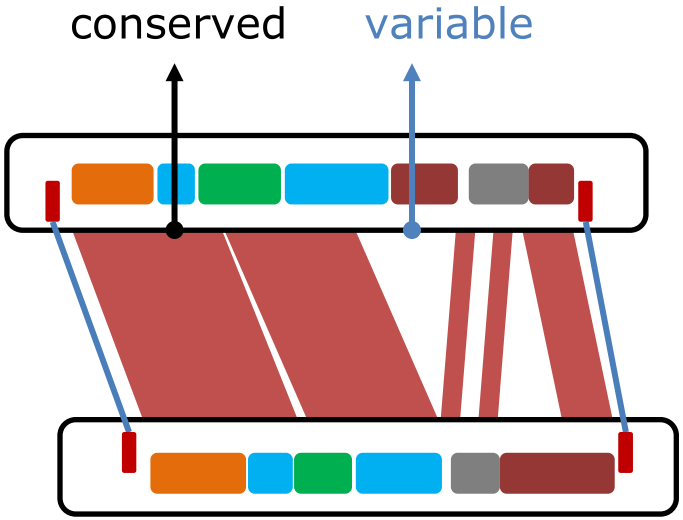
PhD Nathan Papon (2019-2023)

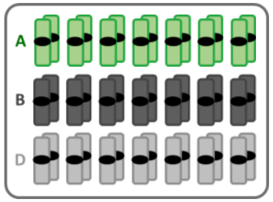






❖ Transposon dynamics





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

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

➢ **Transposition highly regulated; transposons may be s role in shaping a functional genome architecture**

bioRxiv posts many COVID19-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, Hélène Rimbert, Romain De Oliveira, Etienne Paux, Frédéric 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?].

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

AuBi
Auvergne Bioinformatique

○ *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

○ *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

UNIVERSITÉ Clermont Auvergne

INRAE