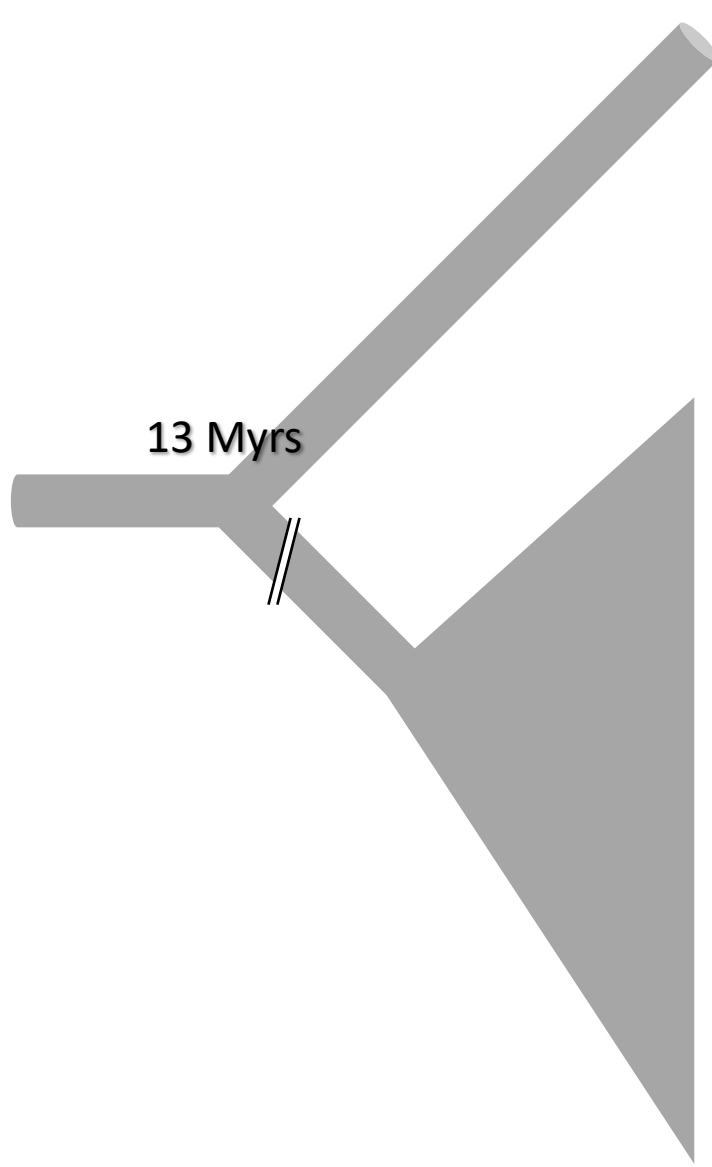


Dynamics of transposable elements in the 16 Gb hexaploid wheat genome

Frédéric Choulet

GDEC, INRA, UCA, Clermont-Ferrand, France



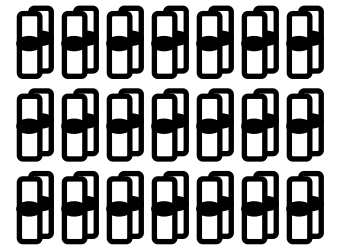


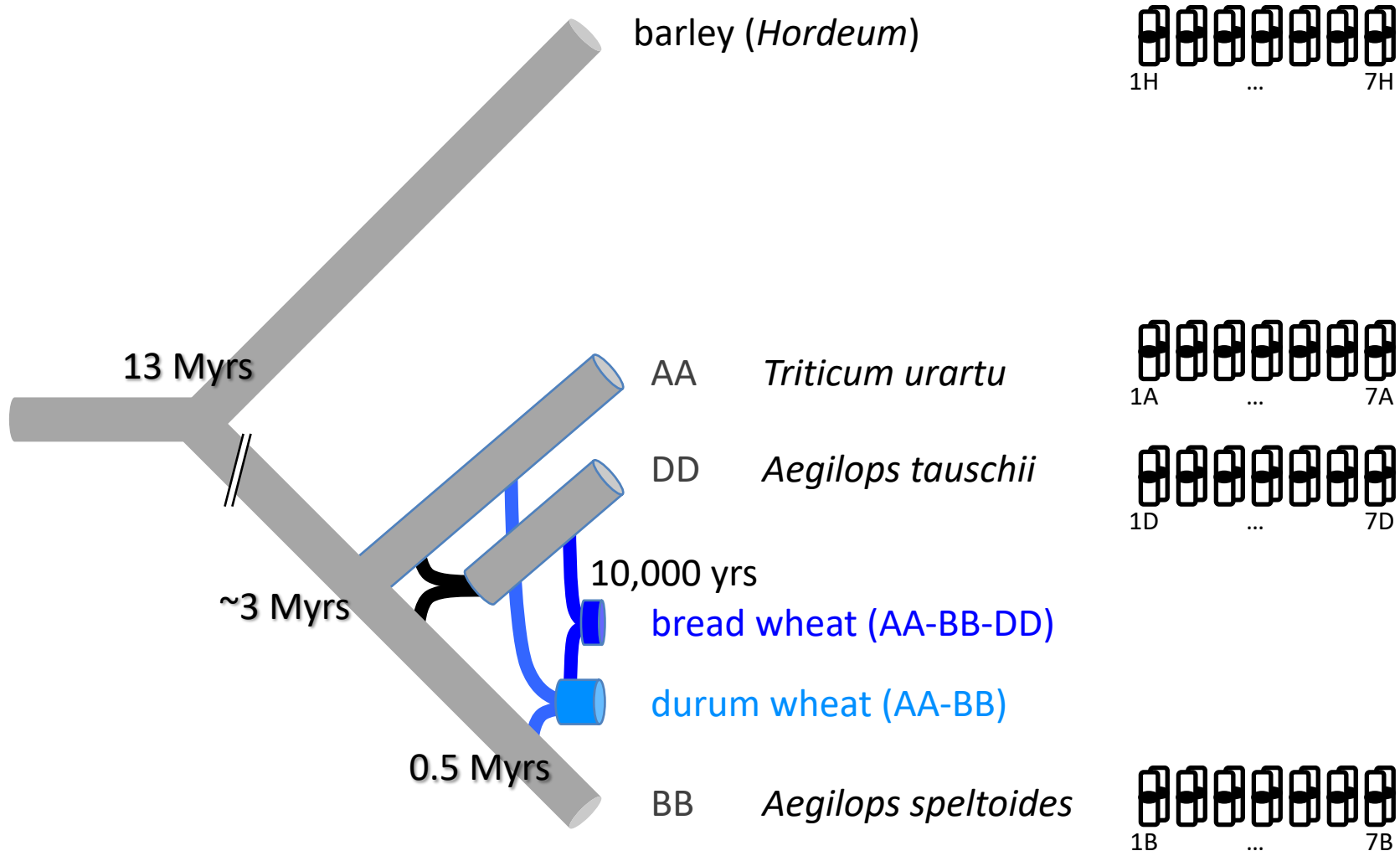
barley (*Hordeum*)

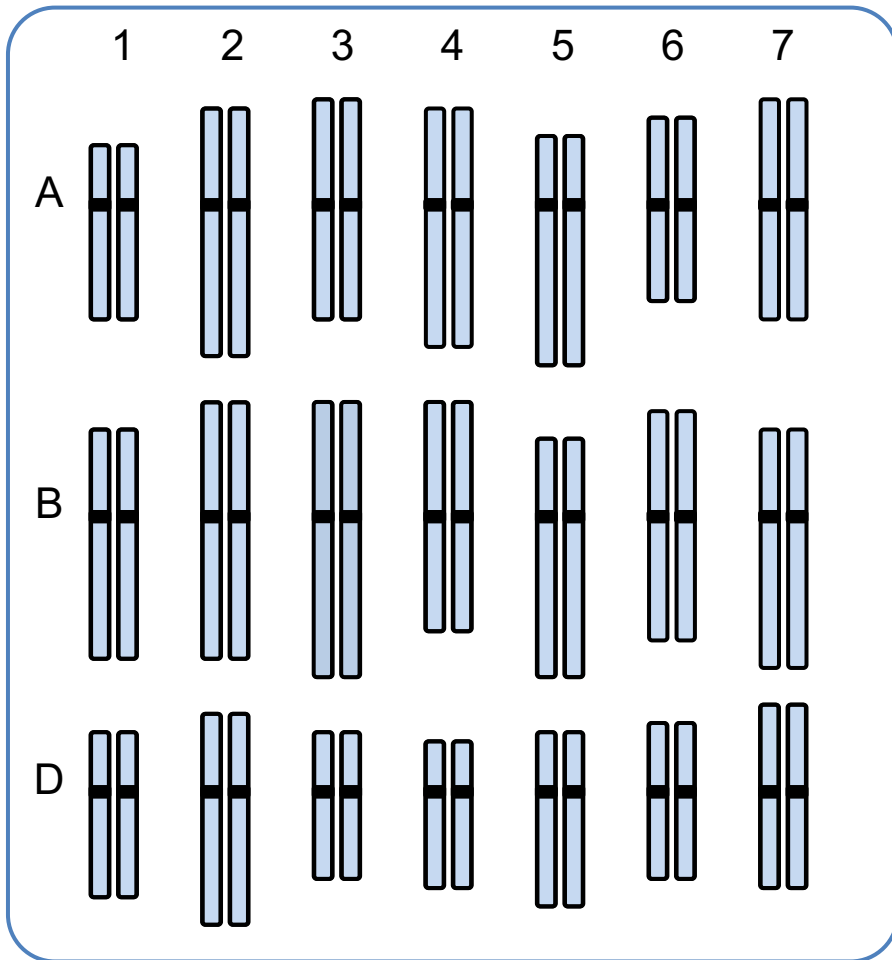


- *Triticum*
- *Aegilops*

bread wheat (AA-BB-DD)
(Triticum aestivum)







Complex genome

- 16 Gb
- Hexaploid AA-BB-DD
- 85% TEs



Launched in 2005

-> Produce a high quality genome sequence of hexaploid wheat





August 17th 2018

RESEARCH

RESEARCH ARTICLE SUMMARY

WHEAT GENOME

Shifting the limits in wheat research and breeding using a fully annotated reference genome

International Wheat Genome Sequencing Consortium (IWGSC)*

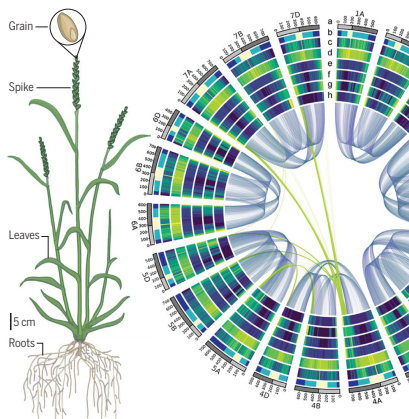
INTRODUCTION: Wheat (*Triticum aestivum* L.) is the most widely cultivated crop on Earth, contributing about a fifth of the total calories consumed by humans. Consequently, wheat yields and production affect the global economy, and failed harvests can lead to social unrest. Breeders continuously strive to develop improved varieties by fine-tuning genetically complex yield and end-use quality parameters while maintaining stable yields and adapting the crop to regionally specific biotic and abiotic stresses.

RATIONALE: Breeding efforts are limited by insufficient knowledge and understanding of

wheat biology and the central agronomic traits. The human population growth need for wheat research to accelerate genetic gain as plant and animal species annotated and ordered genomic regulatory sequence information, has promoted systematic and more timely for the selection and important traits. Wheat is primarily owing to the chromosome that is more

as the human genome, polyploid, and complex, containing more than 85% repetitive DNA. To provide a foundation for improvement through molecular breeding, in 2005, the International Wheat Genome Sequencing Consortium set out to deliver a high-quality annotated reference genome sequence of bread wheat.

RESULTS: An annotated reference genome representing the hexaploid bread wheat genome in the form of 21 chromosomes, like so



Wheat genome deciphered, assembled, and ordered. Seeds, or grains, are important to wheat yields (left panel), but all parts of the plant contribute to crop production. The complete access to the ordered sequence of all 21 wheat chromosomes, the sequences, and the interaction network of expressed genes—all shown here panel) with concentric tracks for diverse aspects of wheat genome composition. Researchers now have the ability to rewrite the story of wheat crop improvement ranges underlying the concentric heatmaps of the right panel are provided

International Wheat Genome Sequencing Consortium (IWGSC), *Science* 361, 661 (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 Wicker^{1†}, Heidrun Gundlach^{2†}, Manuel Spannagl², Cristobal Uauy³, Philippa Borrill³, Ricardo H. Ramirez-González³, Romain De Oliveira⁴, International Wheat Genome Sequencing Consortium⁵, Klaus F. X. Mayer^{2,6}, Etienne Paux⁴ and Frédéric Choulet^{4*}

Abstract

Background: Transposable elements (TEs) are major components of large plant genomes and main drivers of genome evolution. The most recent assembly of hexaploid bread wheat recovered the highly repetitive TE space in an almost complete chromosomal context and enabled a detailed view into the dynamics of TEs in the A, B, and D subgenomes.

Results: The overall TE content is very similar between the A, B, and D subgenomes, although we find no evidence for bursts of TE amplification after the polyploidization events. Despite the near-complete turnover of TEs since the subgenome lineages diverged from a common ancestor, 76% of TE families are still present in similar proportions in each subgenome. Moreover, spacing between syntenic genes is also conserved, even though syntenic TEs have been replaced by new insertions over time, suggesting that distances between genes, but not sequences, are under evolutionary constraints. The TE composition of the immediate gene vicinity differs from the core intergenic regions. We find the same TE families to be enriched or depleted near genes in all three subgenomes. Evaluations at the subfamily level of timed long terminal repeat-retrotransposon insertions highlight the independent evolution of the diploid A, B, and D lineages before polyploidization and cases of concerted proliferation in the AB tetraploid.

Conclusions: Even though the intergenic space is changed by the TE turnover, an unexpected preservation is observed between the A, B, and D subgenomes for features like TE family proportions, gene spacing, and TE enrichment near genes.

Keywords: Transposable elements, Wheat genome, Genome evolution, LTR retrotransposons, Polyploidy, *Triticum aestivum*

Background

Transposable elements (TEs) are ubiquitous components of genomes and one of the major forces driving genome evolution [1]. They are classified into two classes: retrotransposons (class 1), transposing via reverse transcription of their messenger RNA (mRNA), and DNA transposons (class 2), representing all other types of elements [2]. TEs are small genetic units with the ability to make copies of themselves or move around in the genome. They do not encode a function that would allow them to be maintained by selection across generations; rather, their strategy relies

on their autonomous or non-autonomous amplification. TEs are subject to rapid turnover, are the main contributors of intraspecific genomic diversity, and are the main factor explaining genome size variations. Thus, TEs represent the dynamic reservoir of the genomes. They are epigenetically silenced [3], preventing them from long-term massive amplification that could be detrimental. The dynamics of TEs in genomes remains unclear, and it was supposed that they may escape silencing and experience bursts of amplification followed by rapid silencing. Their impact on gene expression has also been documented in many species (for a review, see [4]). In addition, they play a role at the structural level, as essential components of centromeric chromatin in plants [3, 5]. Plant genomes are generally dominated by a small number of highly repeated

* Correspondence: frederic.choulet@inra.fr

†Thomas Wicker and Heidrun Gundlach contributed equally to this work.

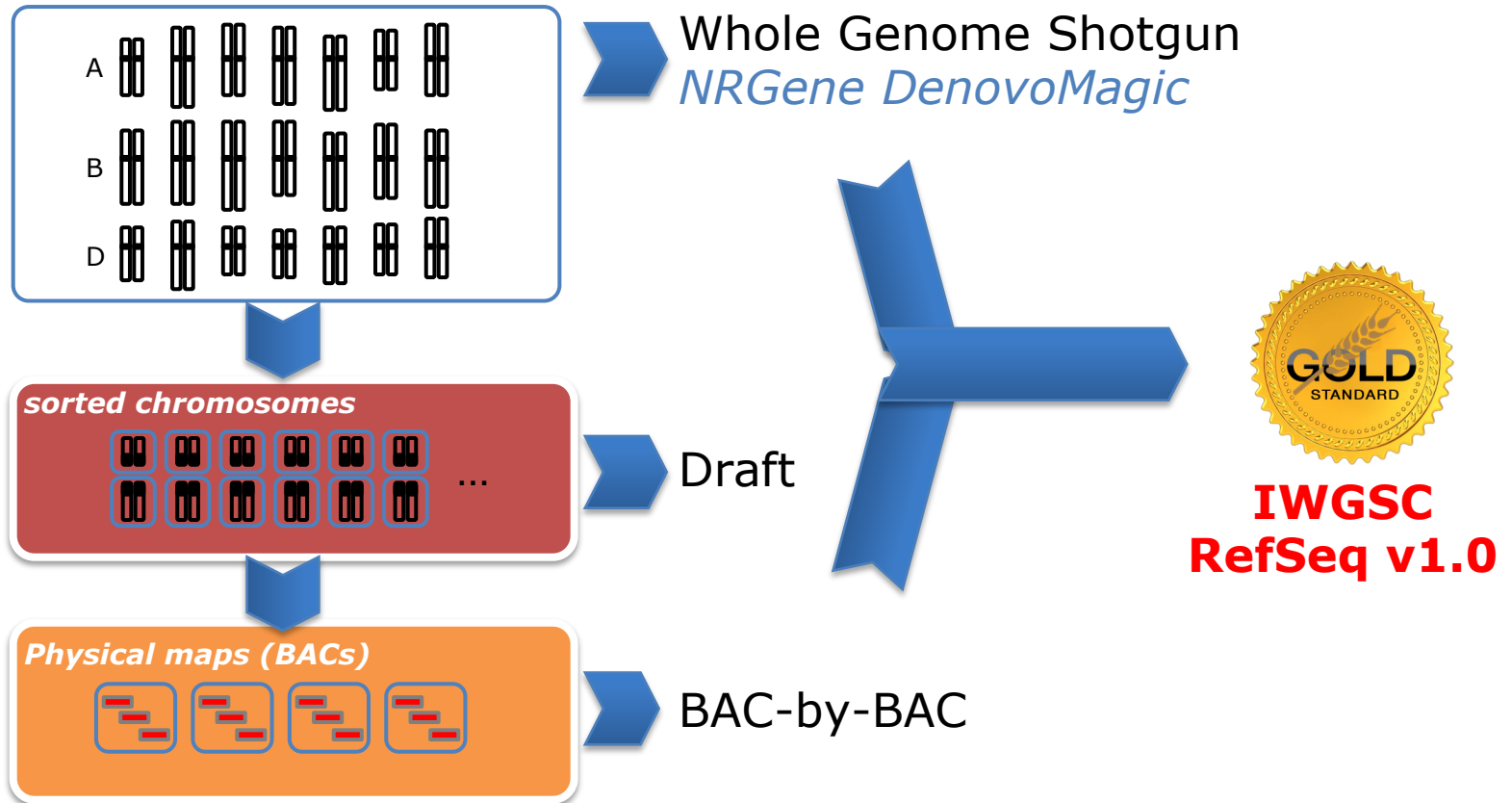
²GDEC, INRA, UCA (Université Clermont Auvergne), Clermont-Ferrand, France
Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

○ IWGSC sequencing strategy

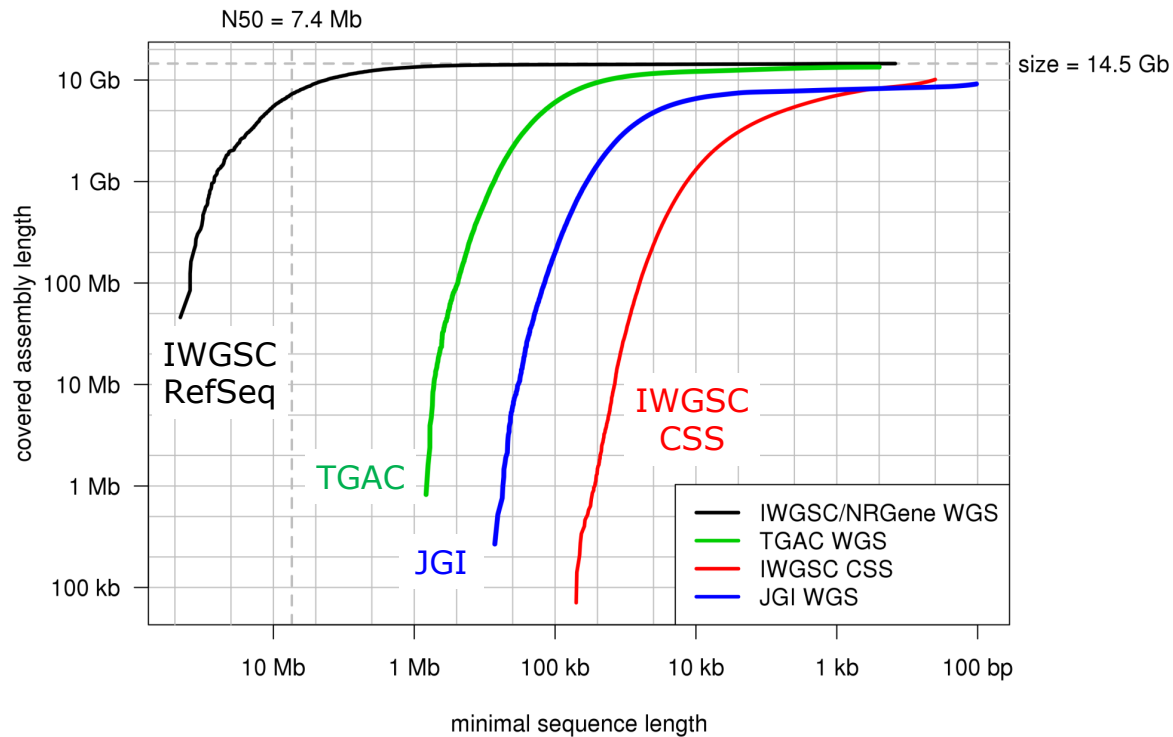
2015 .. 2018



○ IWGSC RefSeq v1.0 – Metrics



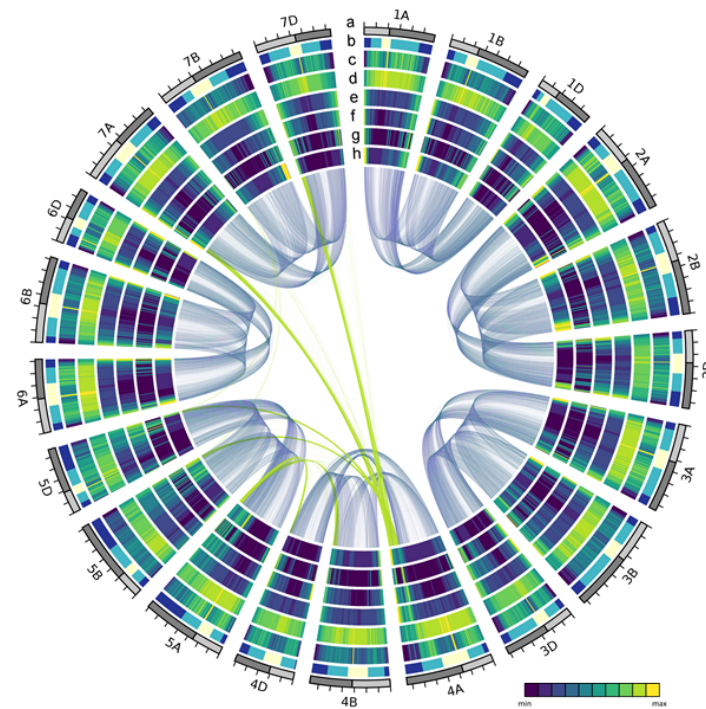
total size 14.5 Gb (21 pseudomolecules = 14.1 Gb)
completeness 97-99%
superscaffold N50 23 Mb



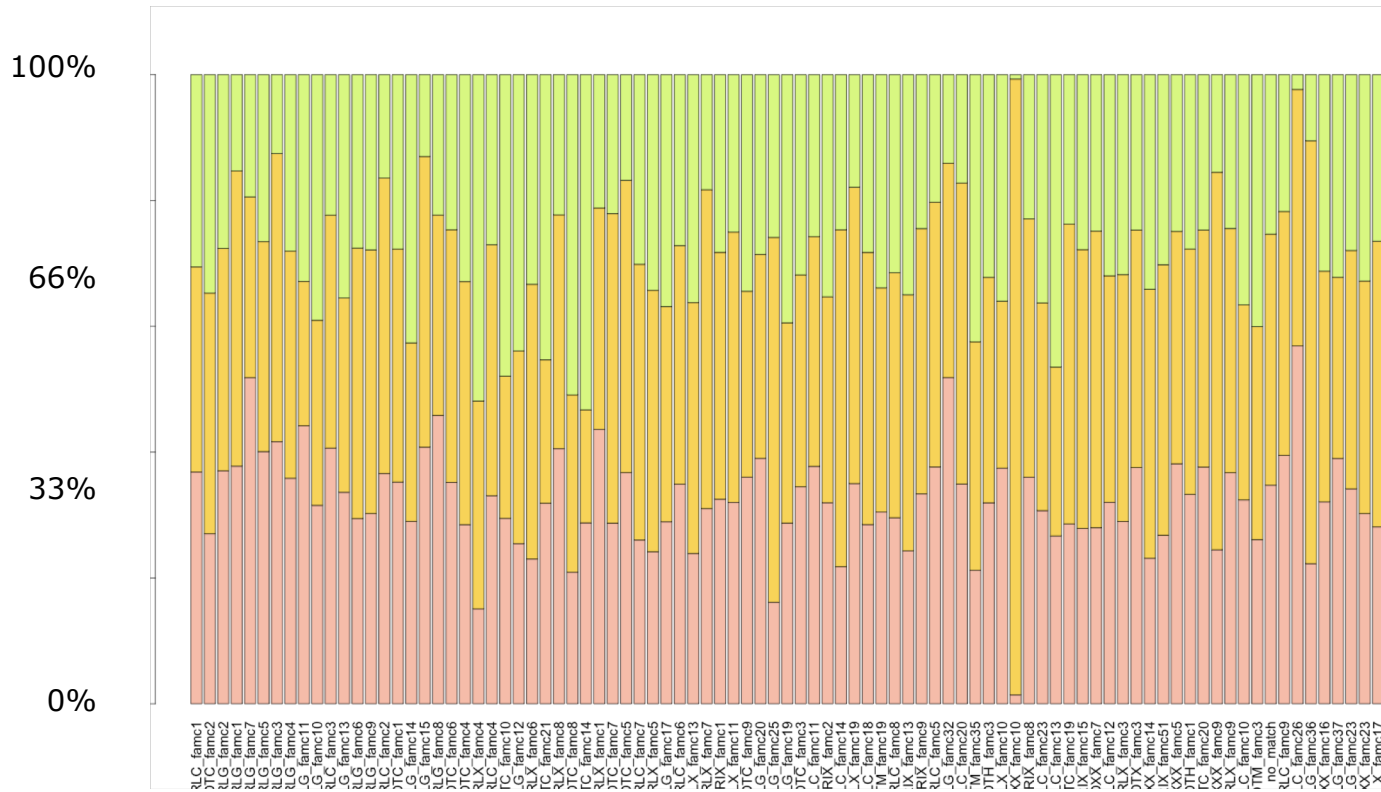
○ IWGSC RefSeq v1.0

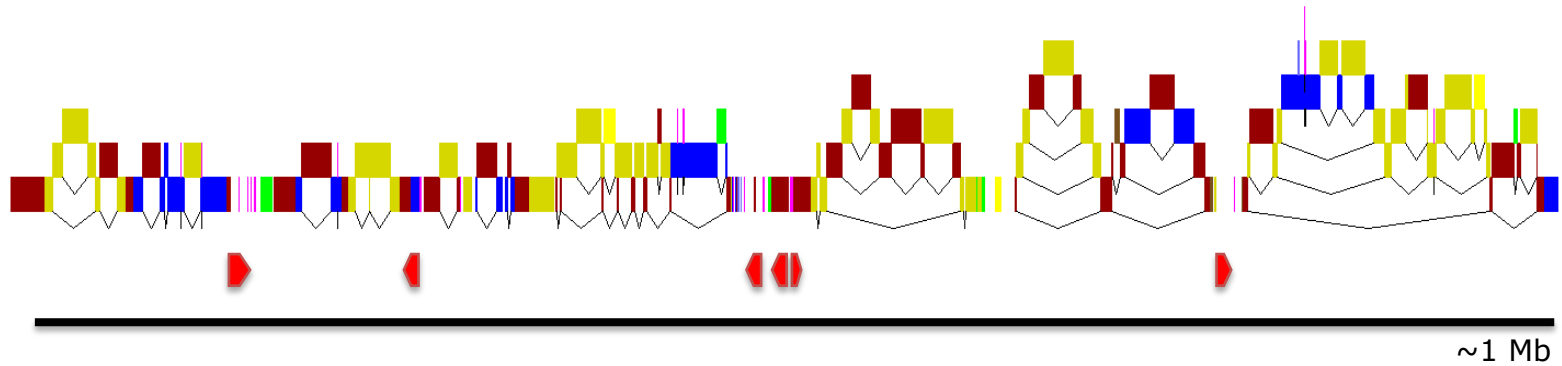
▪ Annotation

- 107,891 genes
- 4 million TEs

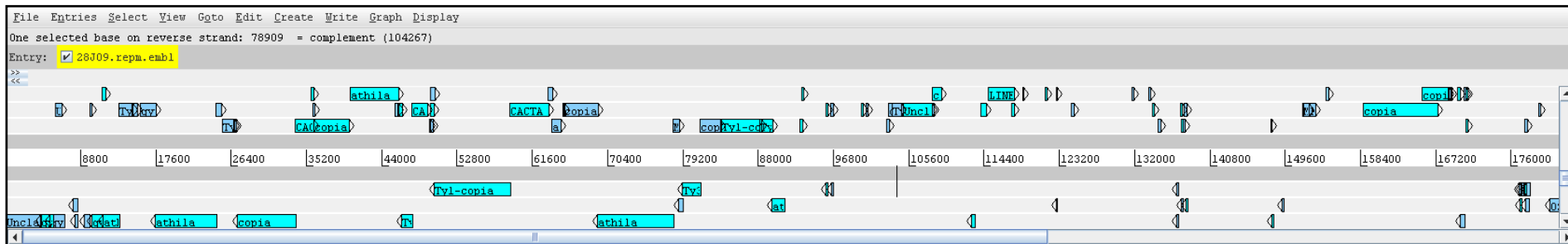


The "1/3 - 1/3 - 1/3" mystery!!!



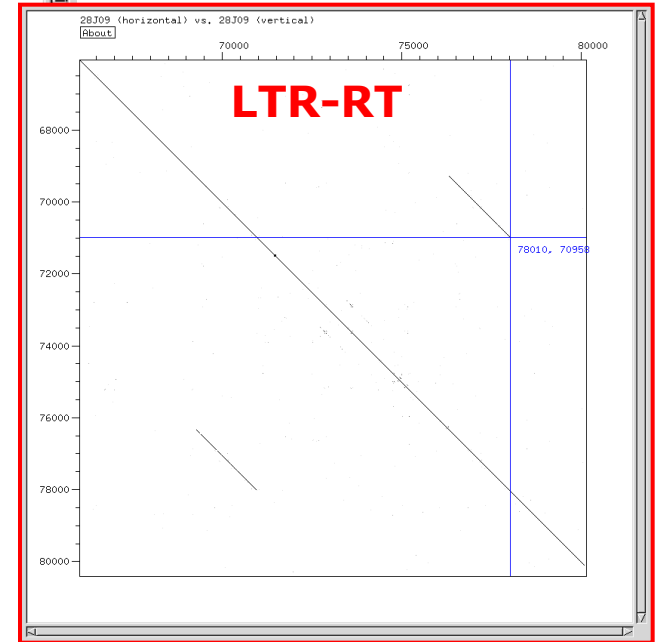
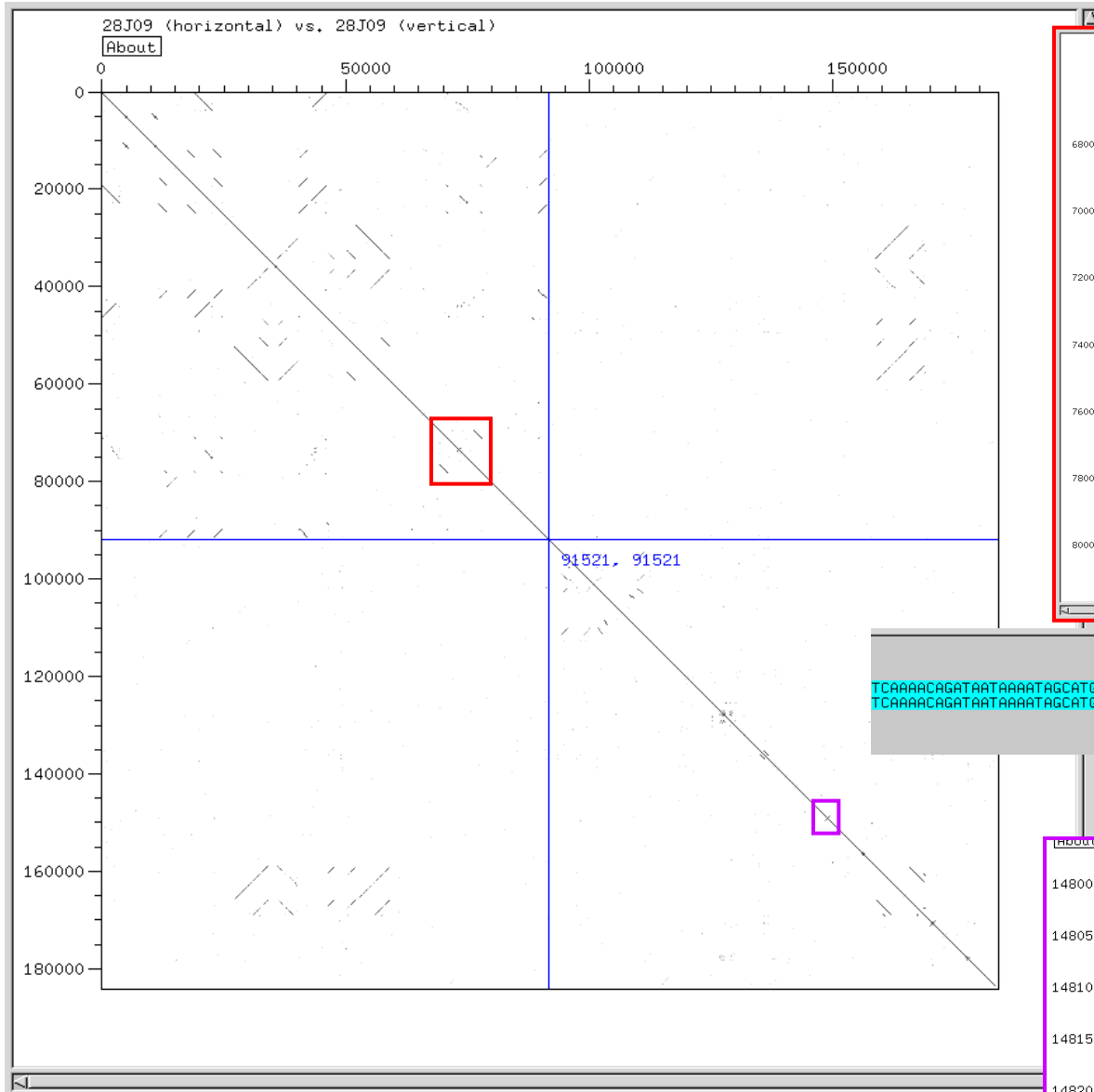


- **TE modeling** based on similarity with a **TE library**
- **Building a reference TE library:**
 - fully de novo
 - curated by experts



105 TE fragments (RepeatMasker with TREP)

Dotter for manual curation

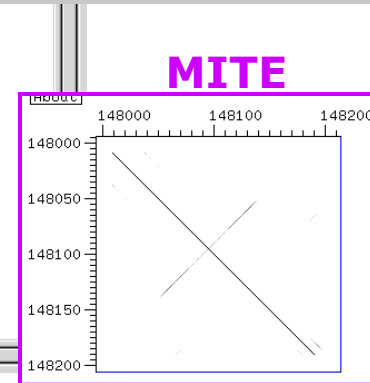


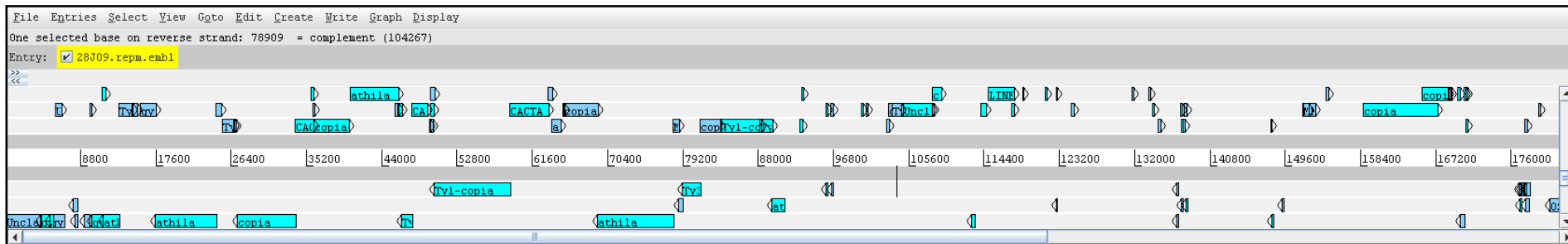
78010

TCARAAACAGATAATAAAATAGCATGAATGCTTCATAAATTATAGATACGTTGGAGACGATATCAGCCTCAATGCACCTGCTCCACT

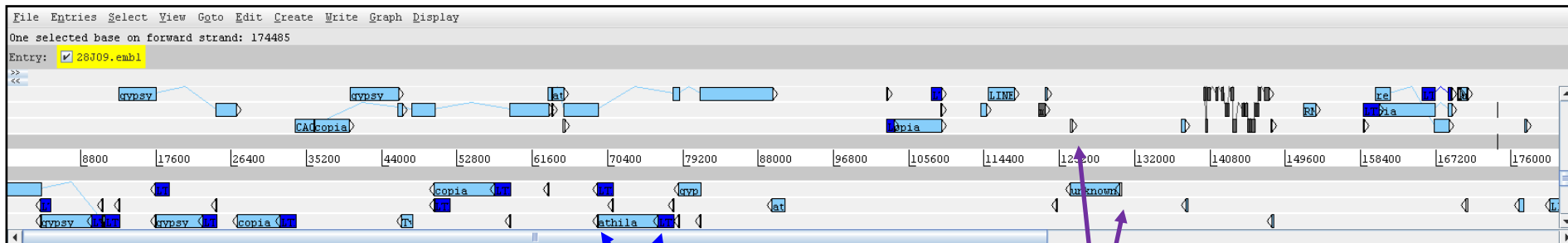
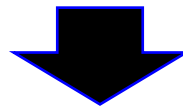
TCARAAACAGATAATAAAATAGCATGAATGCTTCATAAATTATAGATACGTTGGAGACGATATCAGCATCCCAAGCTTAAATCCCTAC

70958





105 TE fragments



32 TEs (manually curated)
+ nested insertions

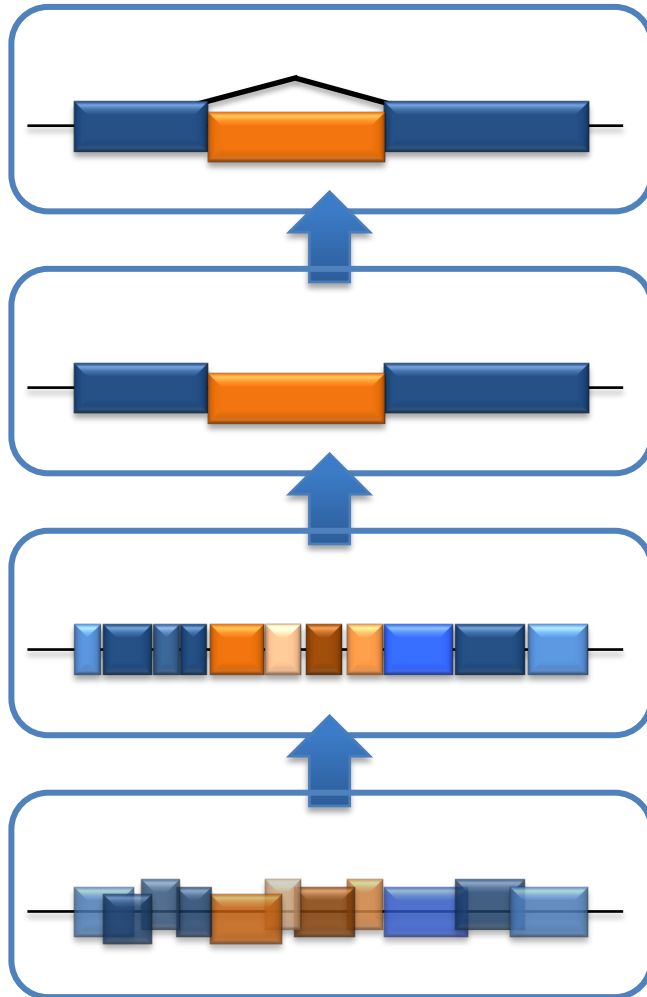
LTRs

TIRs

➤ Manual curation:
Absolutely necessary to have a high quality TE library

○ TE modeling with CLARI-TE and ClariTeRep

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

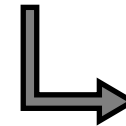


TREP (T. Wicker)
1700 TEs

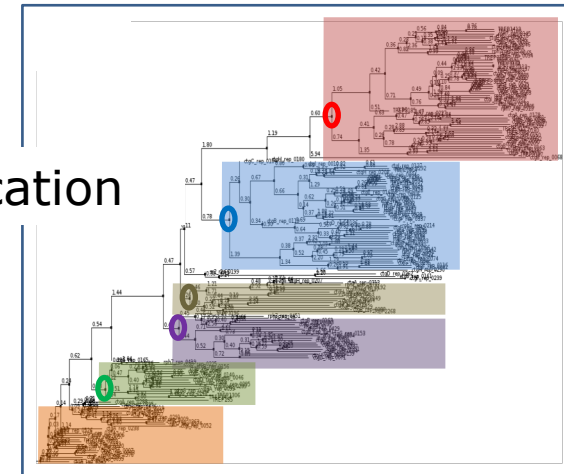
+



3200 TEs
Choulet et al. Plant Cell 2010



Classification

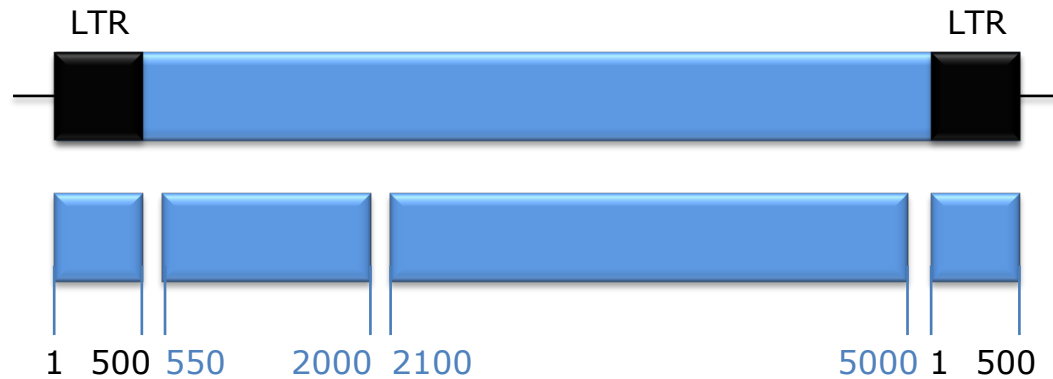


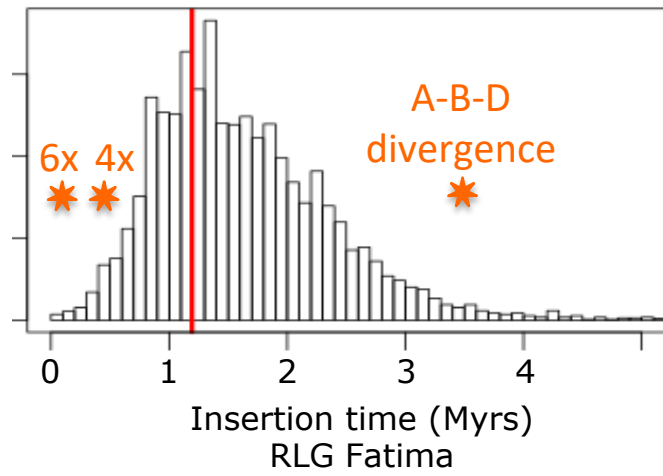
ClariTeRep
~500 families

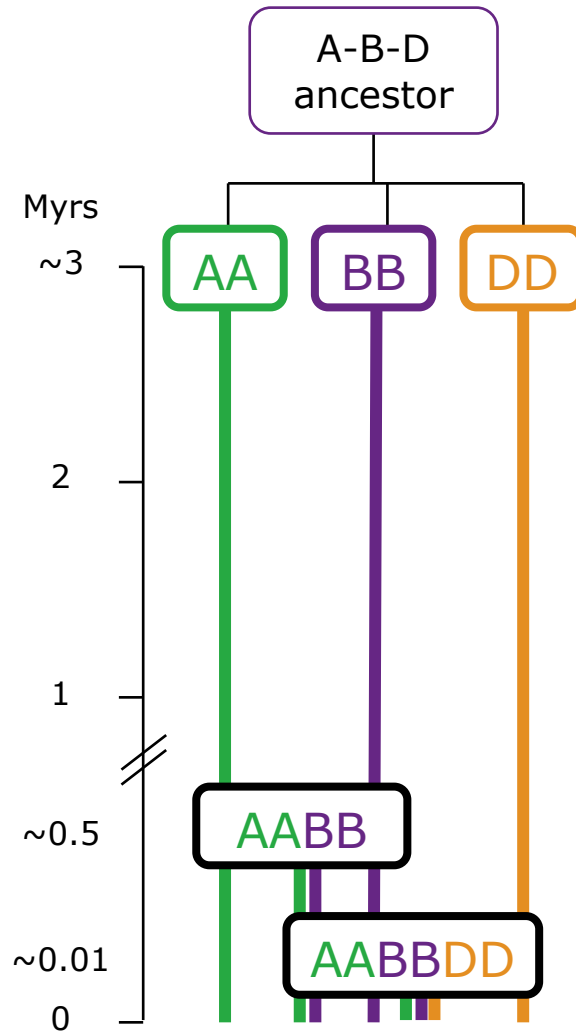


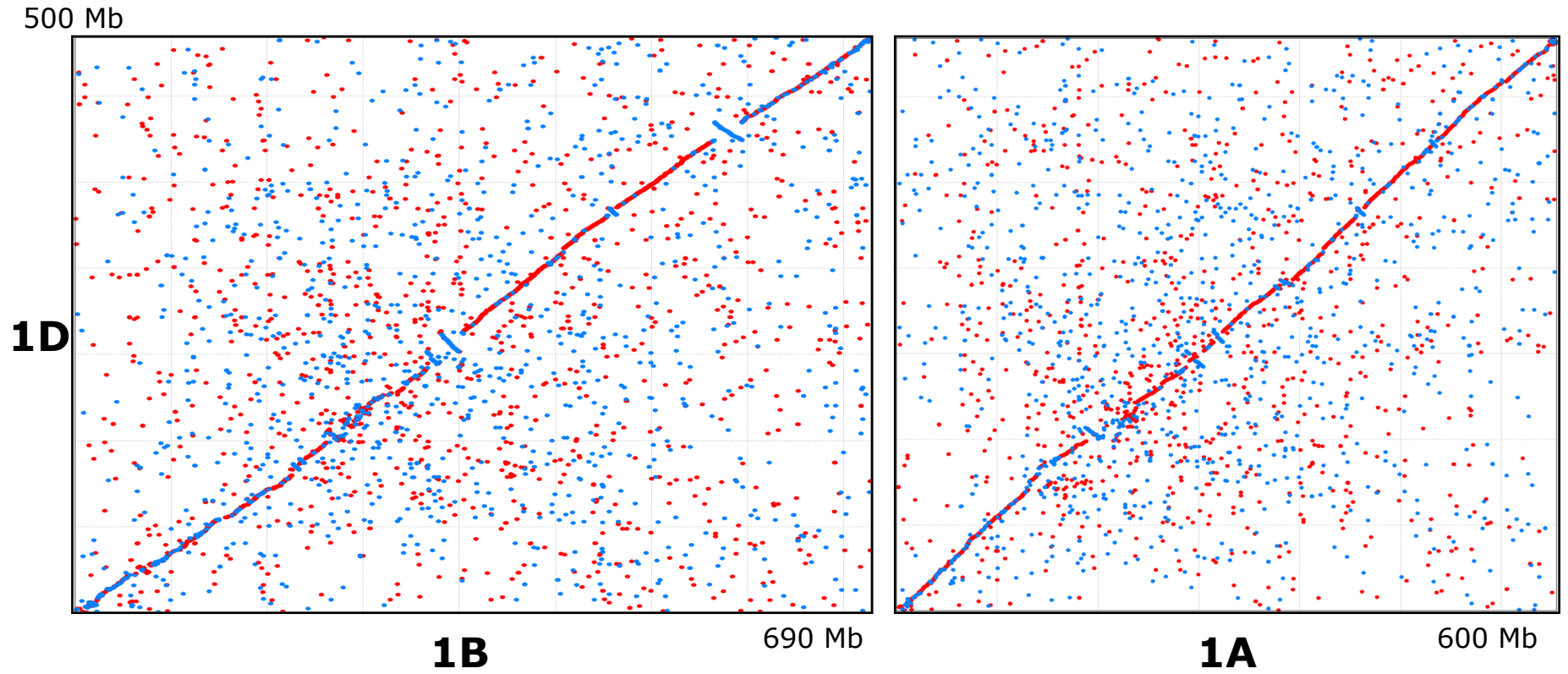
○ TE modeling with CLARI-TE and ClariTeRep

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

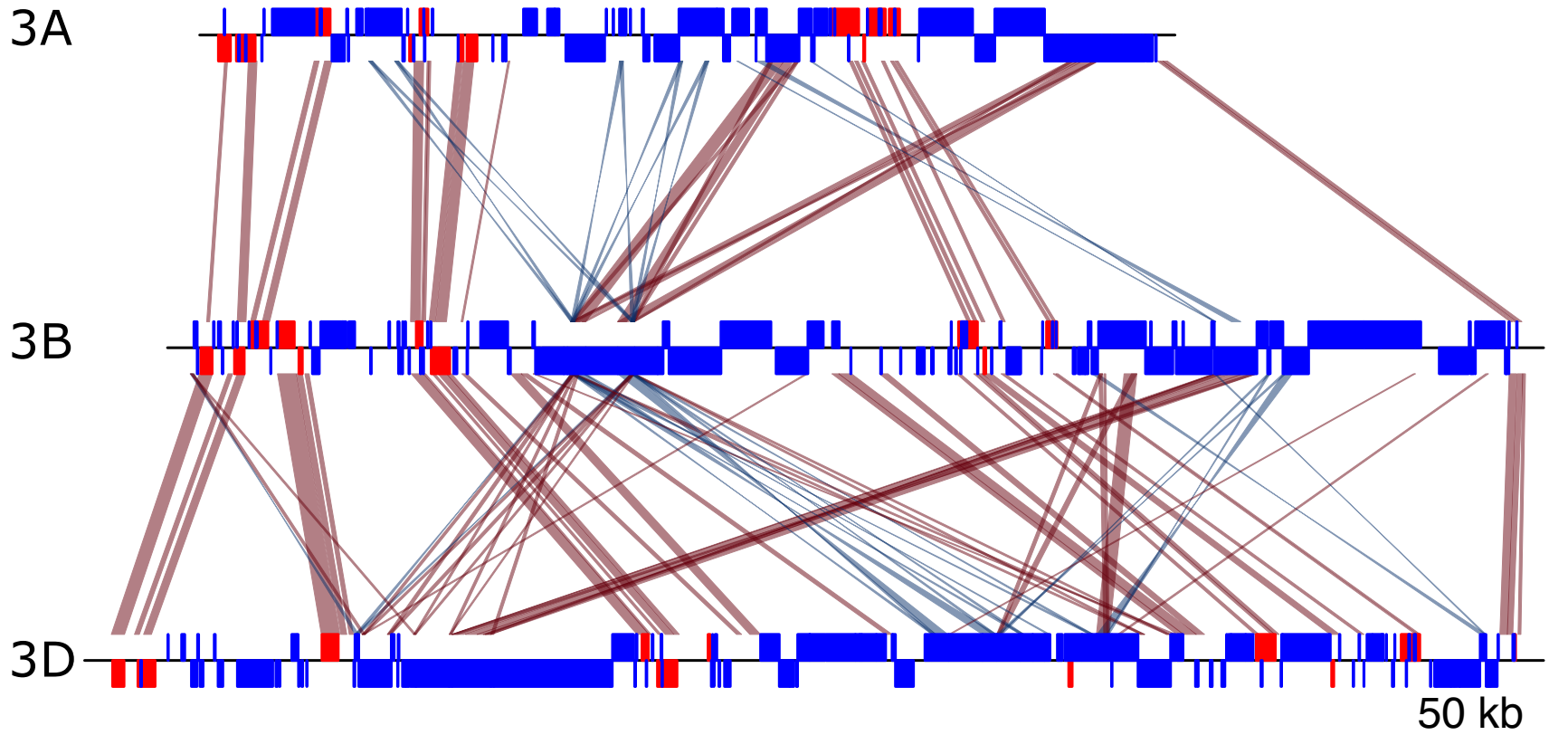




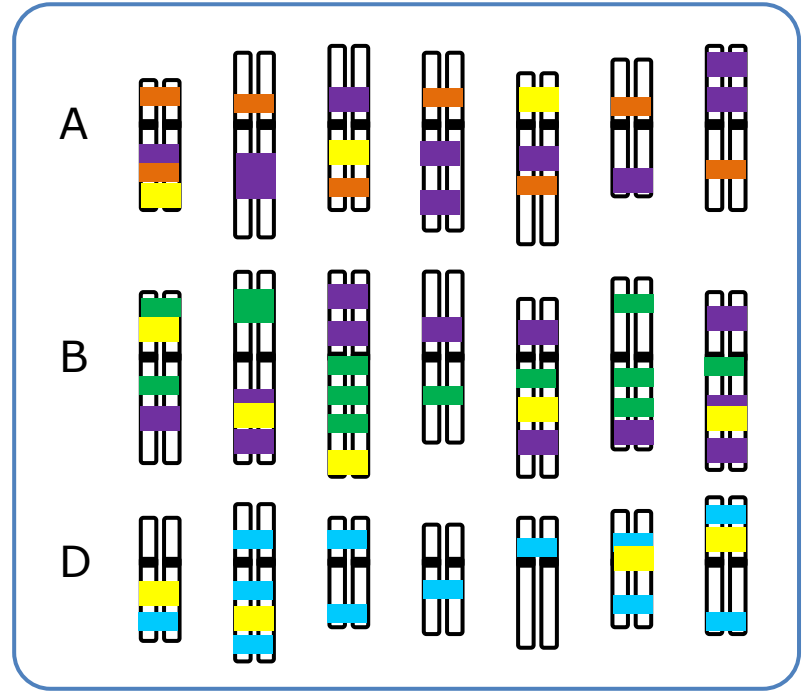
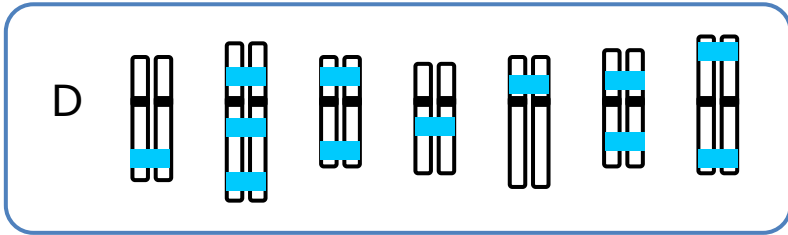
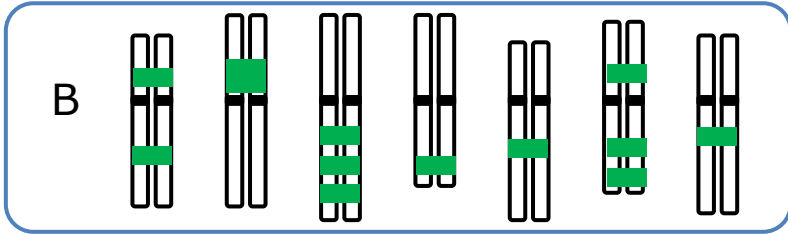
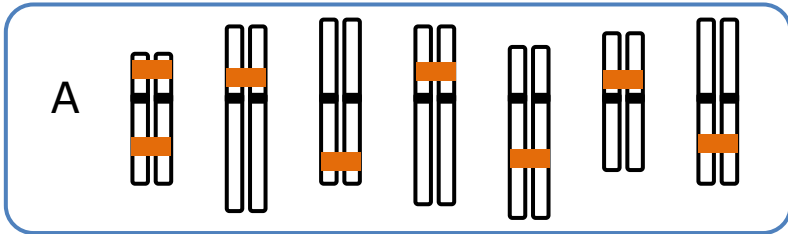




➤ High level of synteny between A-B-D (gene-based)

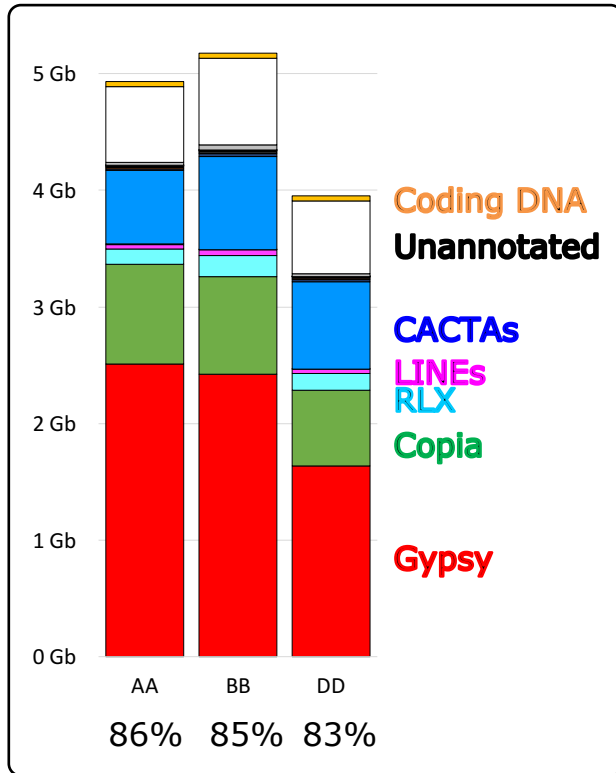


➤ Near-complete **TE turnover** since A-B-D divergence

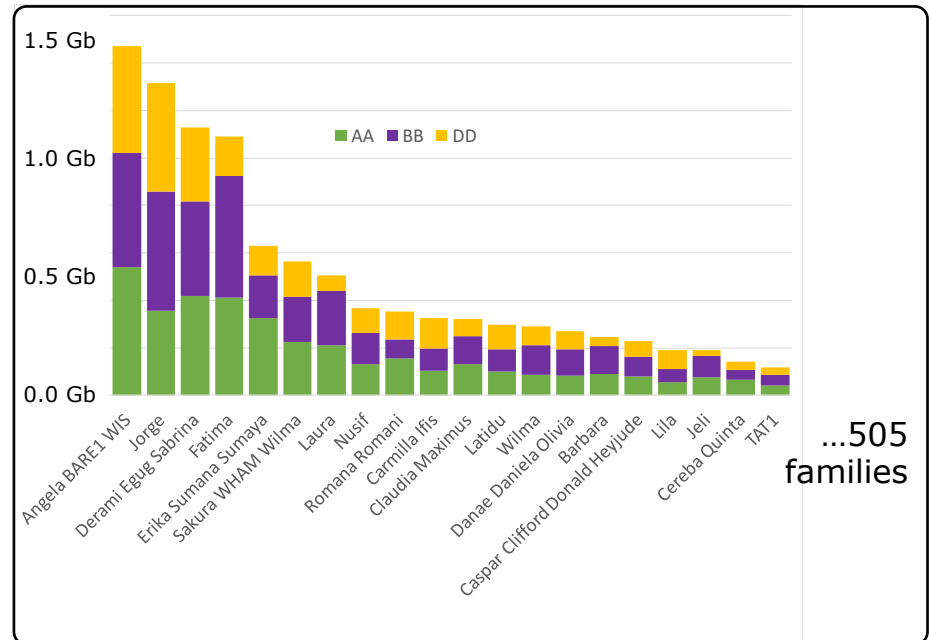


○ % TEs

Superfamilies



Families



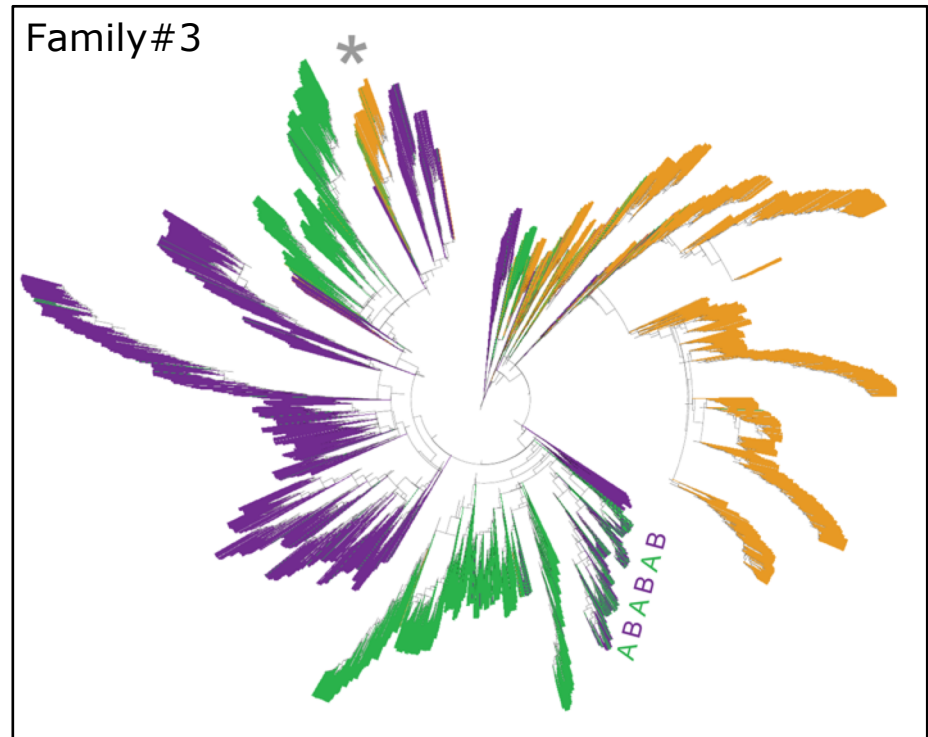
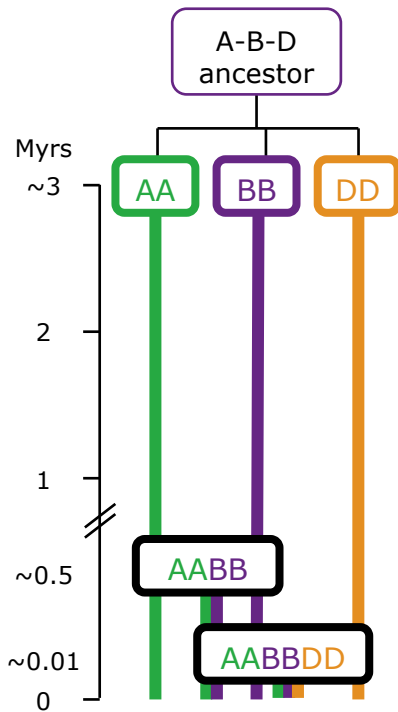
➤ Proportion of each superfamily is very similar betw A-B-D

➤ Abundant families are the same on A-B-D

➤ **0** family specific to 1 subgenome

➤ **76%** of the fam. account for similar % on A-B-D (<2 fold-change)

○ TE trees



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

○ TE dynamics

- TE fam conserved betw A-B-D
→ ancestrally present
- TE turnover BUT proportions remained similar
→ most families were active
→ "equilibrium" between deletion / amplification

- Hypothesis:

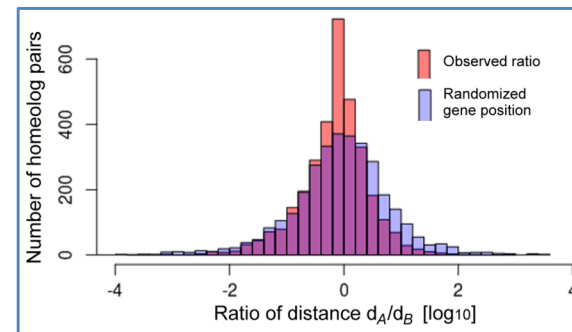
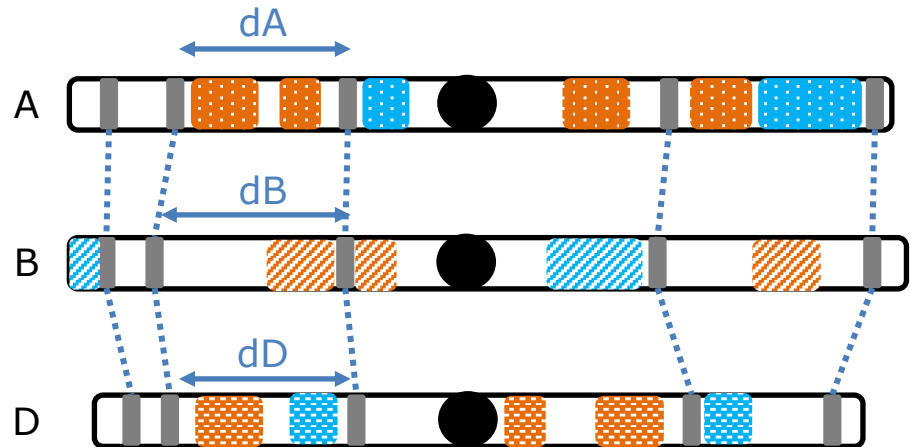
Structural role of TEs likely under selection pressure?

→ Intergenic distances tend to be conserved!!! ... genome architecture likely under evolutionary constraints

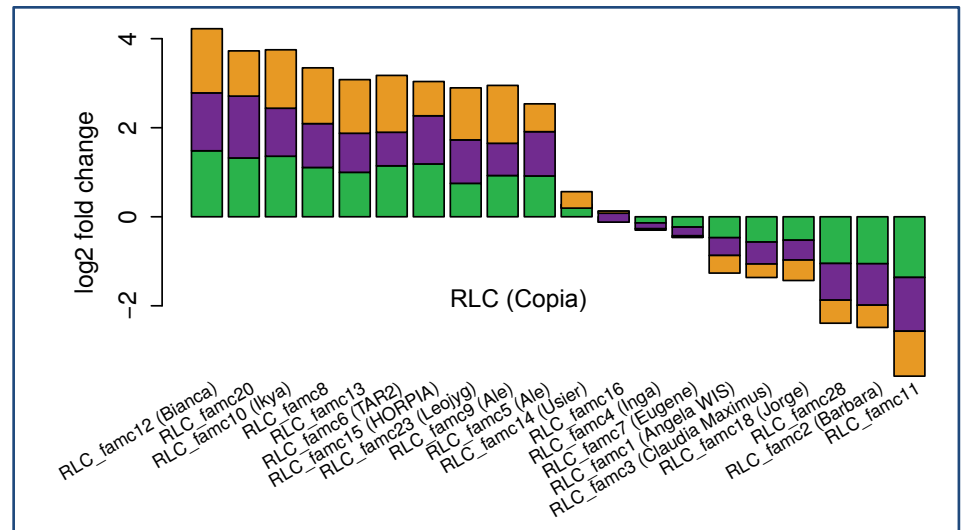
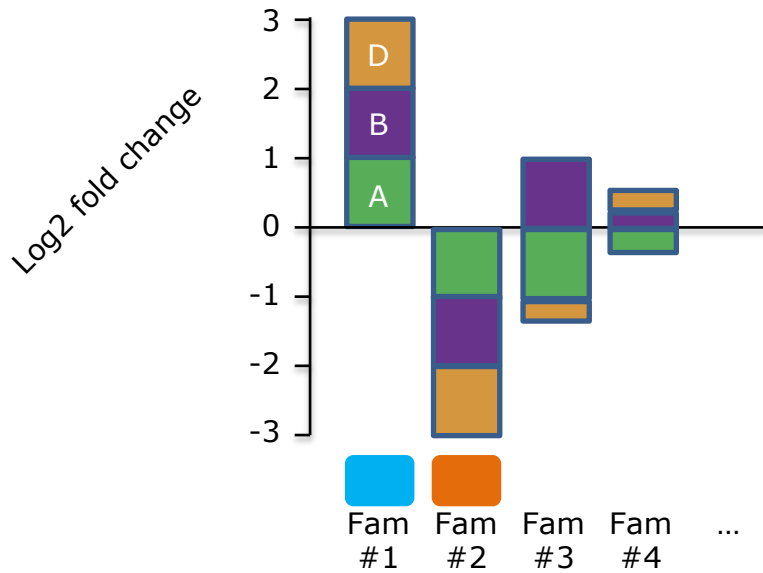
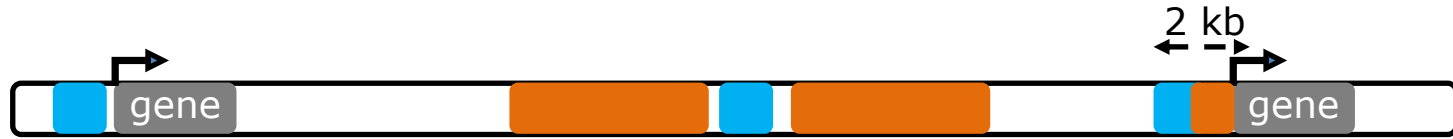
A-B-D last common ancestor



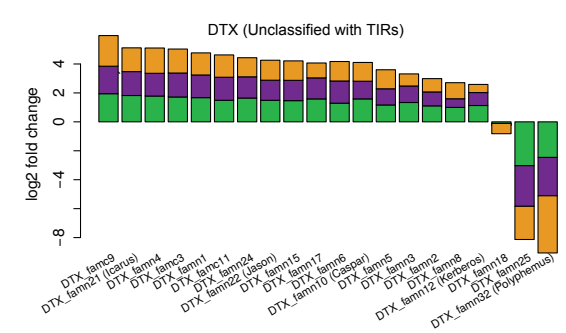
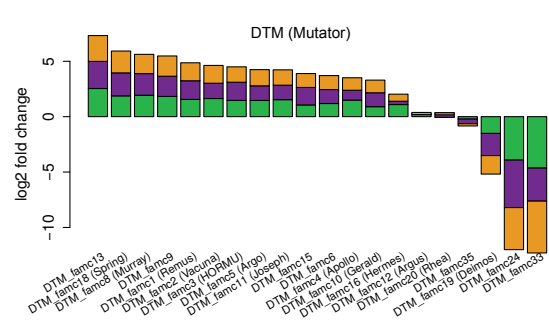
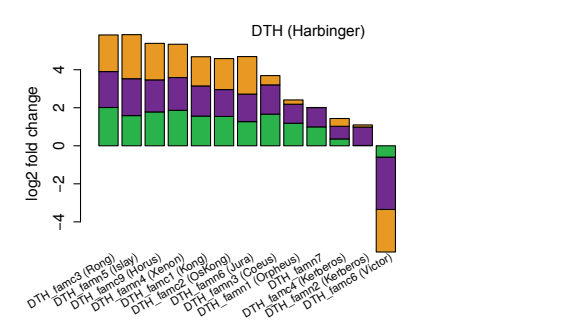
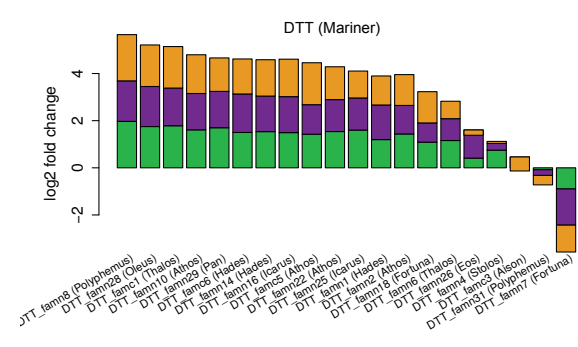
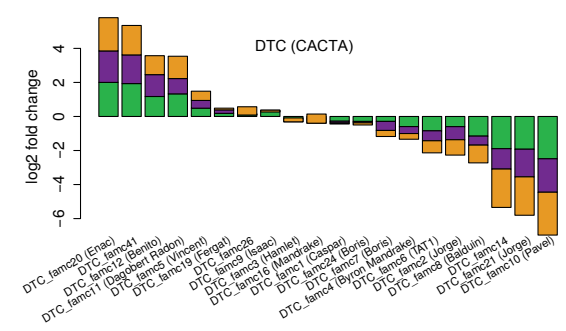
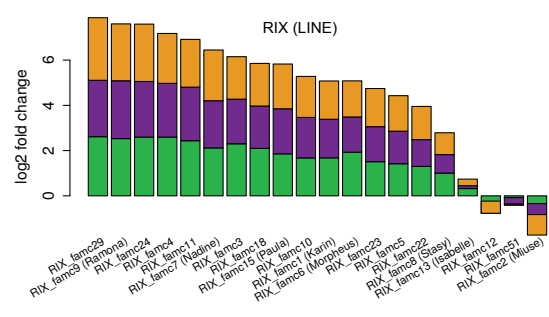
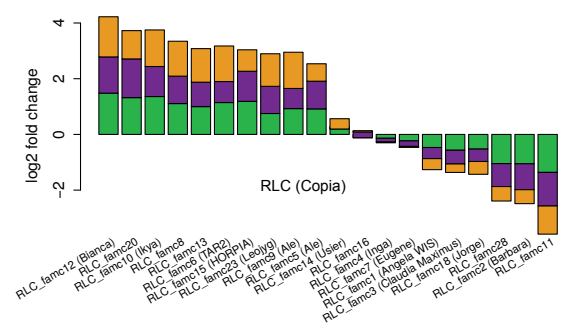
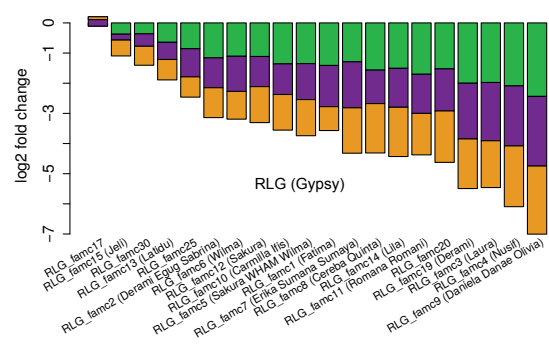
~ 3 Myrs
TE turnover



○ TE content around genes



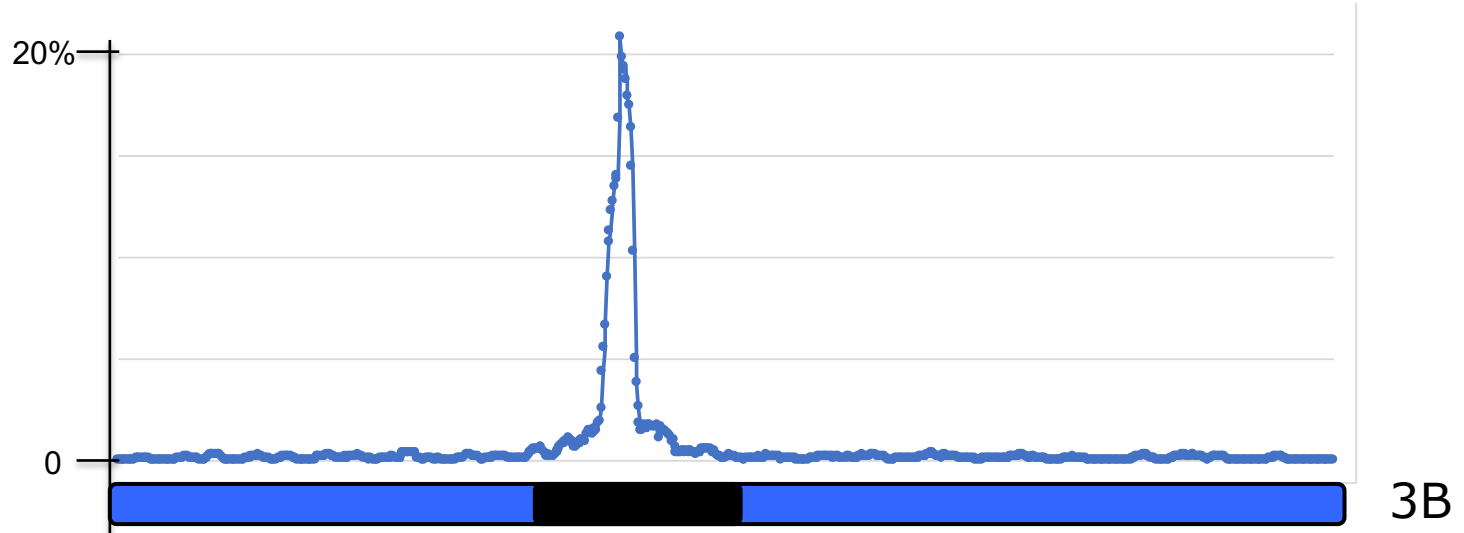
- Wide majority of TE families either enriched/underrepres. in gene promoters
- Enrichment/Underrepres. close to genes is conserved betw A-B-D



○ **Conclusions**

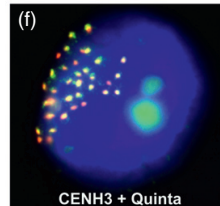
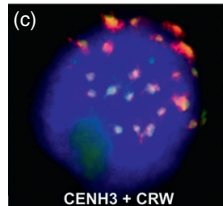
- Nearly-complete TE turnover since A-B-D divergence
- Stability+++ (TE %, families, enrichment around genes...)

=> Hypothesis of a structural role of TEs likely under selection pressure



Cereba

Quinta



Li et al. Plant Journal 2012

→ Structural role of TEs = CenH3-binding site

□ Acknowledgments

○ *University of Zurich*

Thomas Wicker

○ *PGSB, Munich*

Heidrun Gundlach

Manuel Spannagl

Klaus Mayer

○ *John Innes Center, Norwich*

Cristobal Uauy

Philippa Borrill

Ricardo Ramírez-González

○ *GDEC, INRA, UCA, Clermont-Fd*

Romain De Oliveira

Hélène Rimbart

Philippe Leroy

Nicolas Guilhot

Josquin Daron

Ambre A. Josselin

Etienne Paux



○ *IWGSC RefSeq*

R. Appels, K. Eversole,

J. Rogers, N. Stein

...

~200 contributors