Traces of past transposable element presence in Brassicaceae genome dark matter

Hadi Quesneville

Hadi Quesneville. Traces of past transposable element presence in Brassicaceae genome dark matter. ALPHY 2019: Bioinformatics and Evolutionary Genomics, Feb 2019, Paris, France. hal-02734925

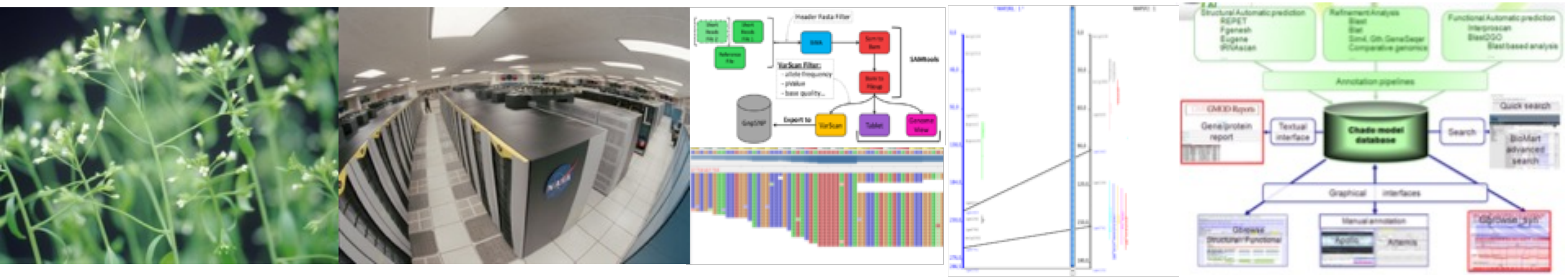
HAL Id: hal-02734925
https://hal.inrae.fr/hal-02734925

Submitted on 2 Jun 2020

Traces of past transposable element presence in *Brassicaceae* genome dark matter

Hadi Quesneville

Alphy, 7 February 2019



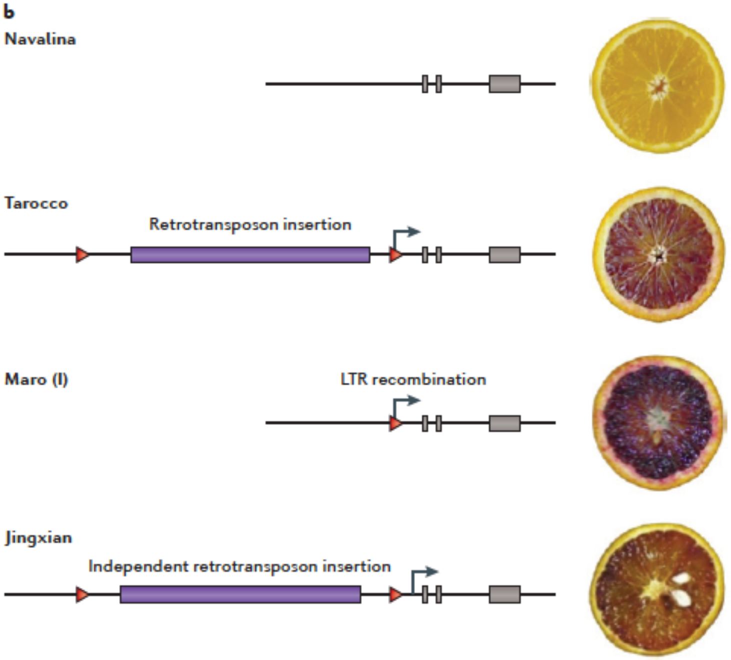
ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

TEs are key players of genome evolution

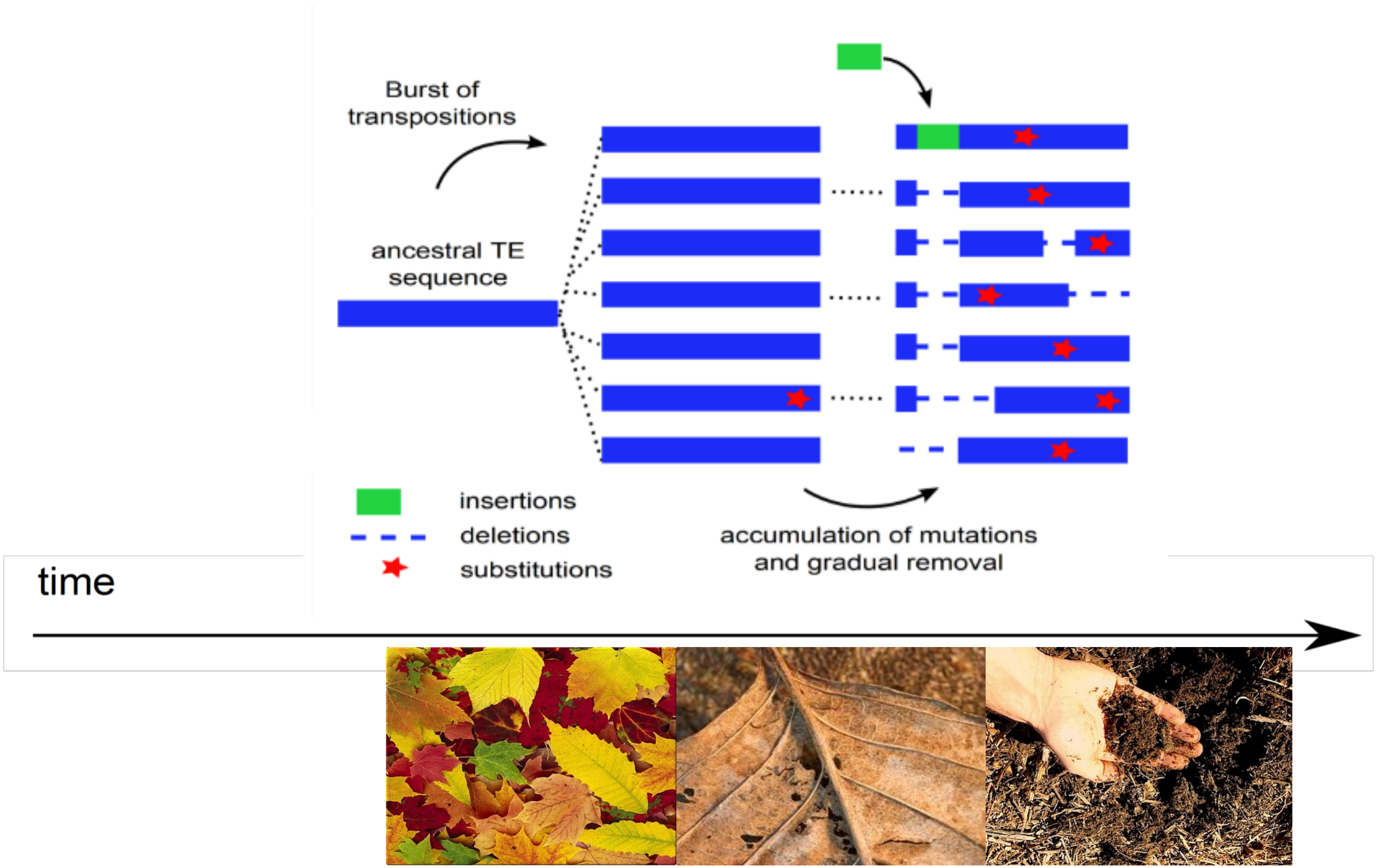
How important are transposons for plant evolution?

Damon Lisch

NATURE REVIEWS | GENETICS VOLUME 14 | JANUARY 2013 | 59



TE dynamics



Hadi Quesneville

Scientific questions

TEs participate to the DNA turnover forming the raw material for genetic innovations.

→ How to recognize very old and degenerated repeated sequences up to 40 Myr old.

→ Search whether they played key evolutionary roles.

Cross-species TE annotation

nature COMMUNICATIONS

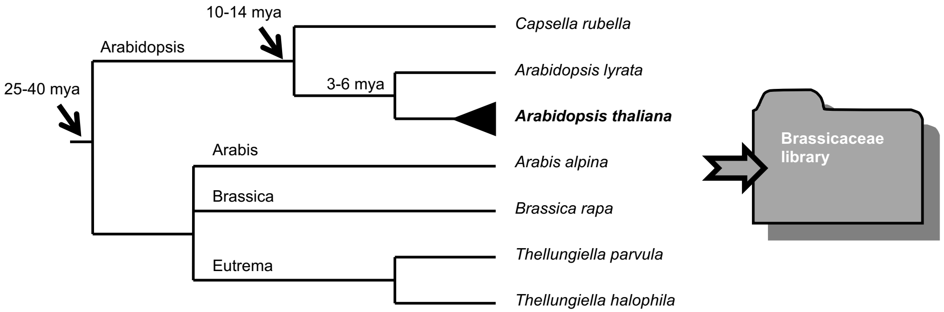
ARTICLE

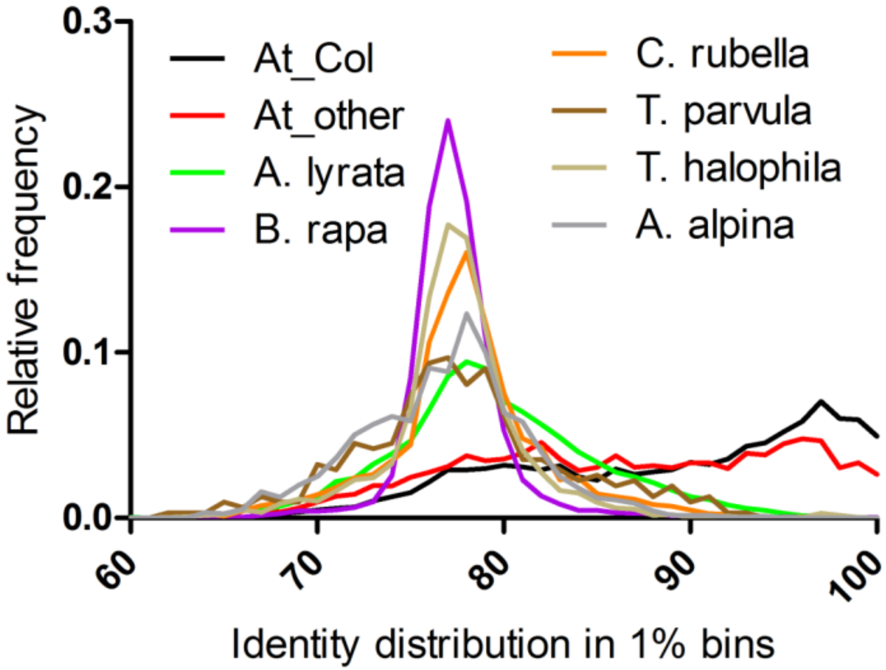
Received 2 Nov 2013 | Accepted 13 May 2014 | Published 23 Jun 2014 | Updated 27 Jun 2014

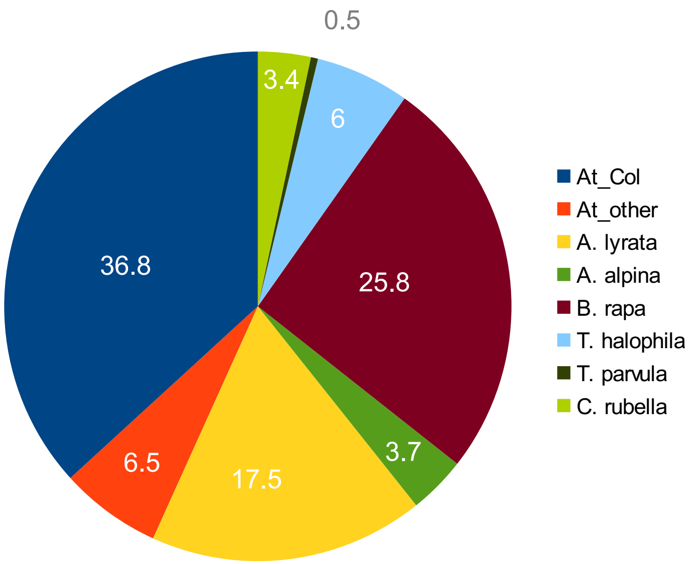
DOI: 10.1038/ncomms5104

Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*

Florian Maumus[1] & Hadi Quesneville[1]







Hadi Quesneville

OPEN ACCESS Freely available online

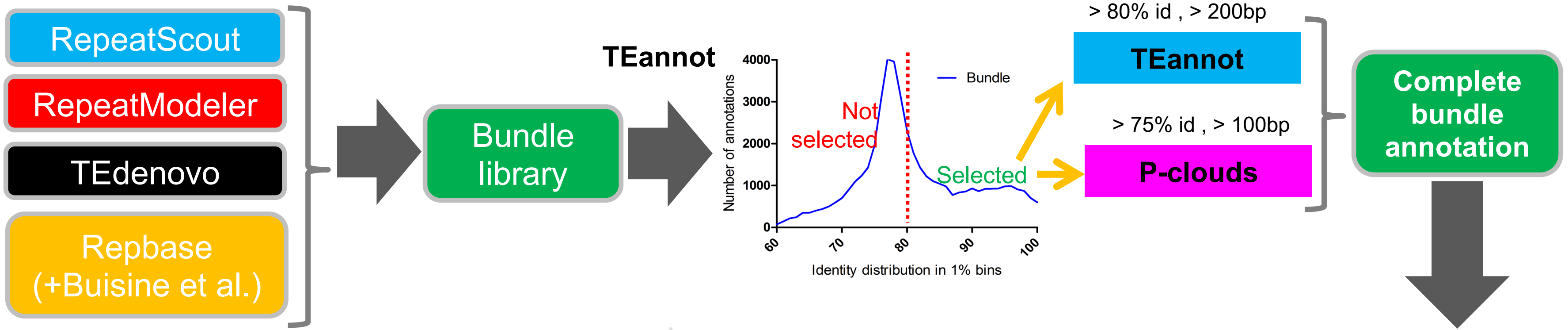
PLOS | ONE

Deep Investigation of *Arabidopsis thaliana* Junk DNA Reveals a Continuum between Repetitive Elements and Genomic Dark Matter

2014

Florian Maumus*, Hadi Quesneville*

UR1164 URGI - Research Unit in Genomics-Info, INRA, Versailles, France


RepeatScout

RepeatModeler

TEdenovo

Repbase
(+Buisine et al.)

Bundle library

TEannot

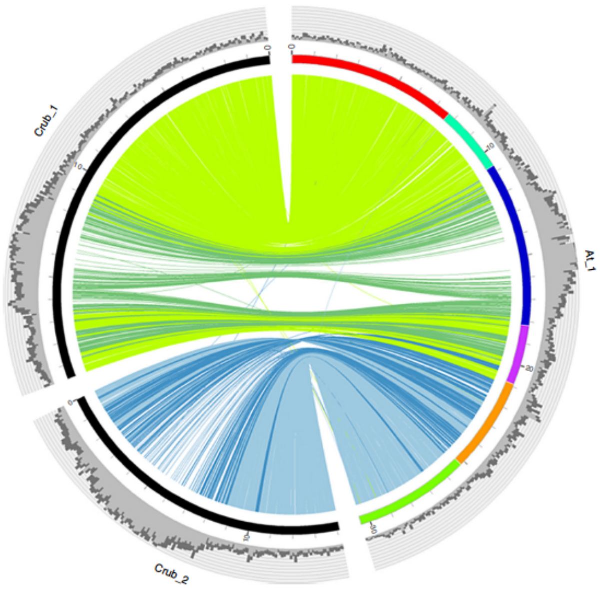
> 80% id , > 200bp

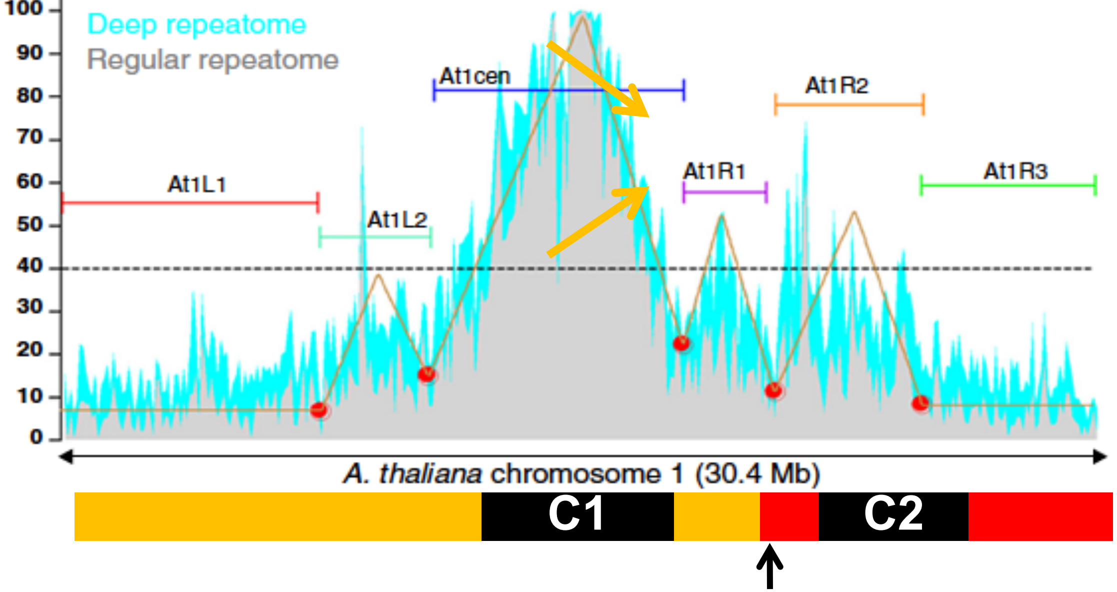
TEannot

> 75% id , > 100bp

P-clouds

Complete bundle annotation





Fusion of two ancestral chromosomes

Hadi Quesneville

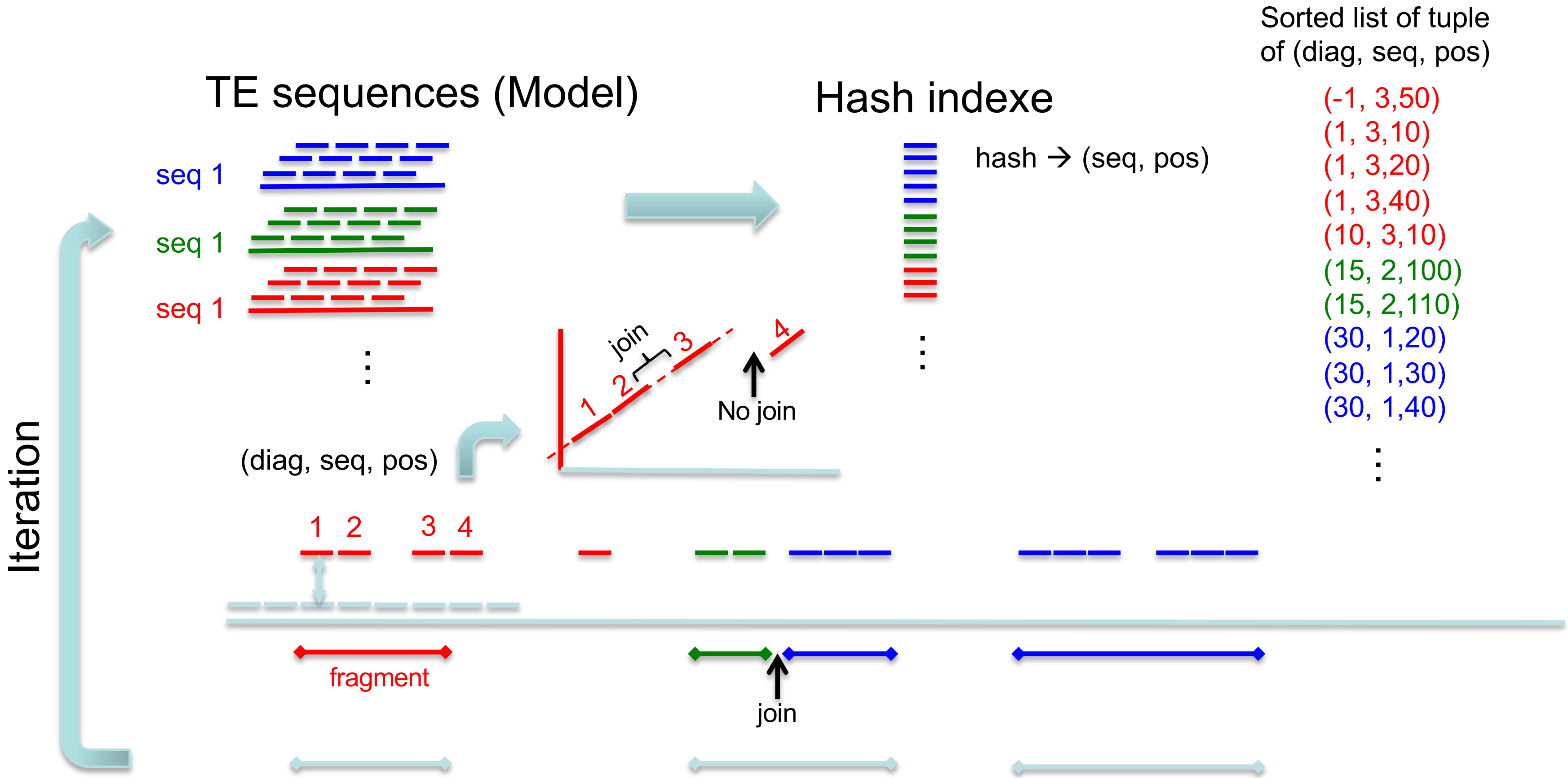
Mixed strategy

- Use TE copies instead of TE consensus for annotation
- Use annotation from many related species to take advantage of « cross-species annotation strategy »

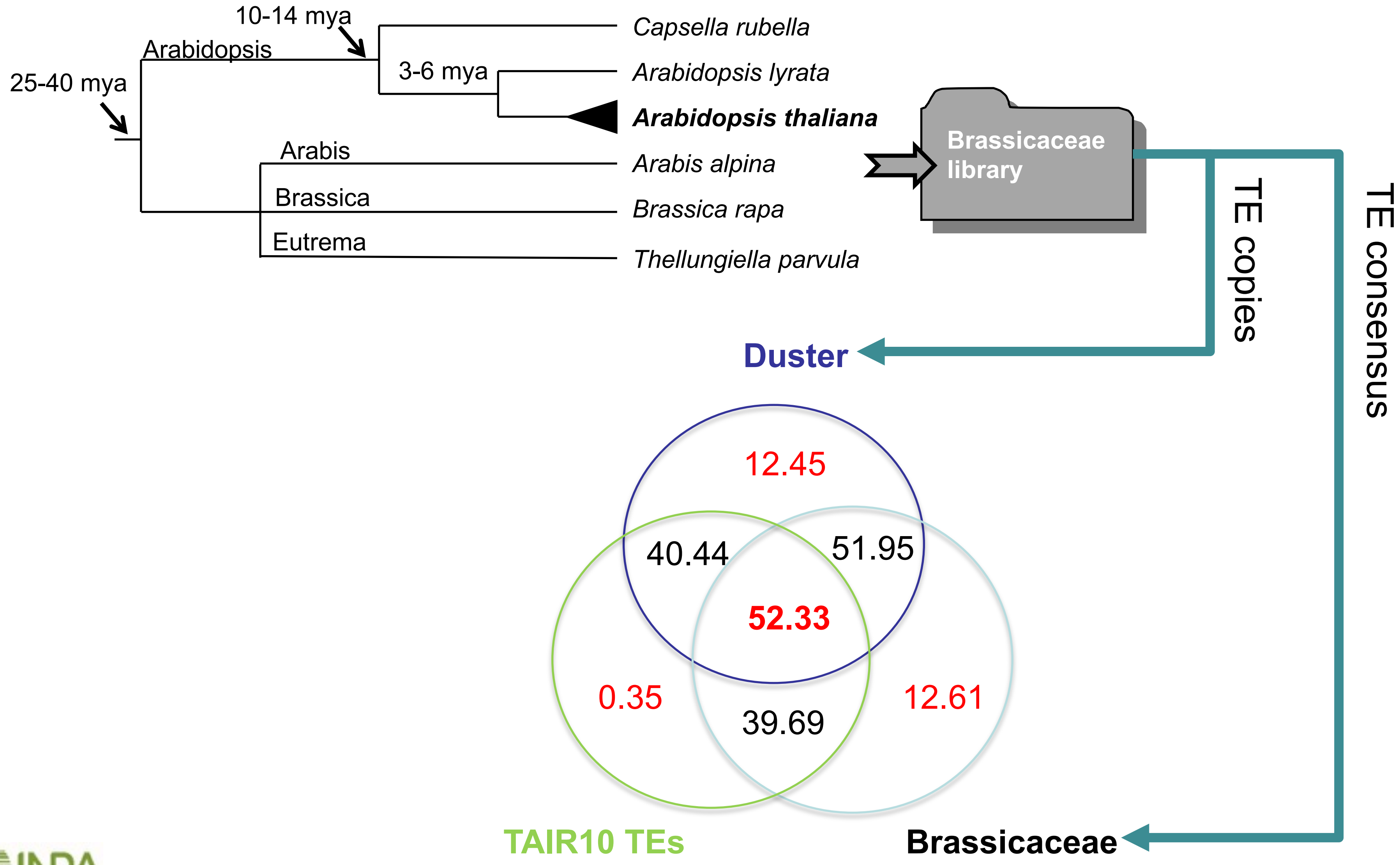
→ Need efficient algorithm for massive comparisons

Duster algorithm

k-mer with regularely spaced holes

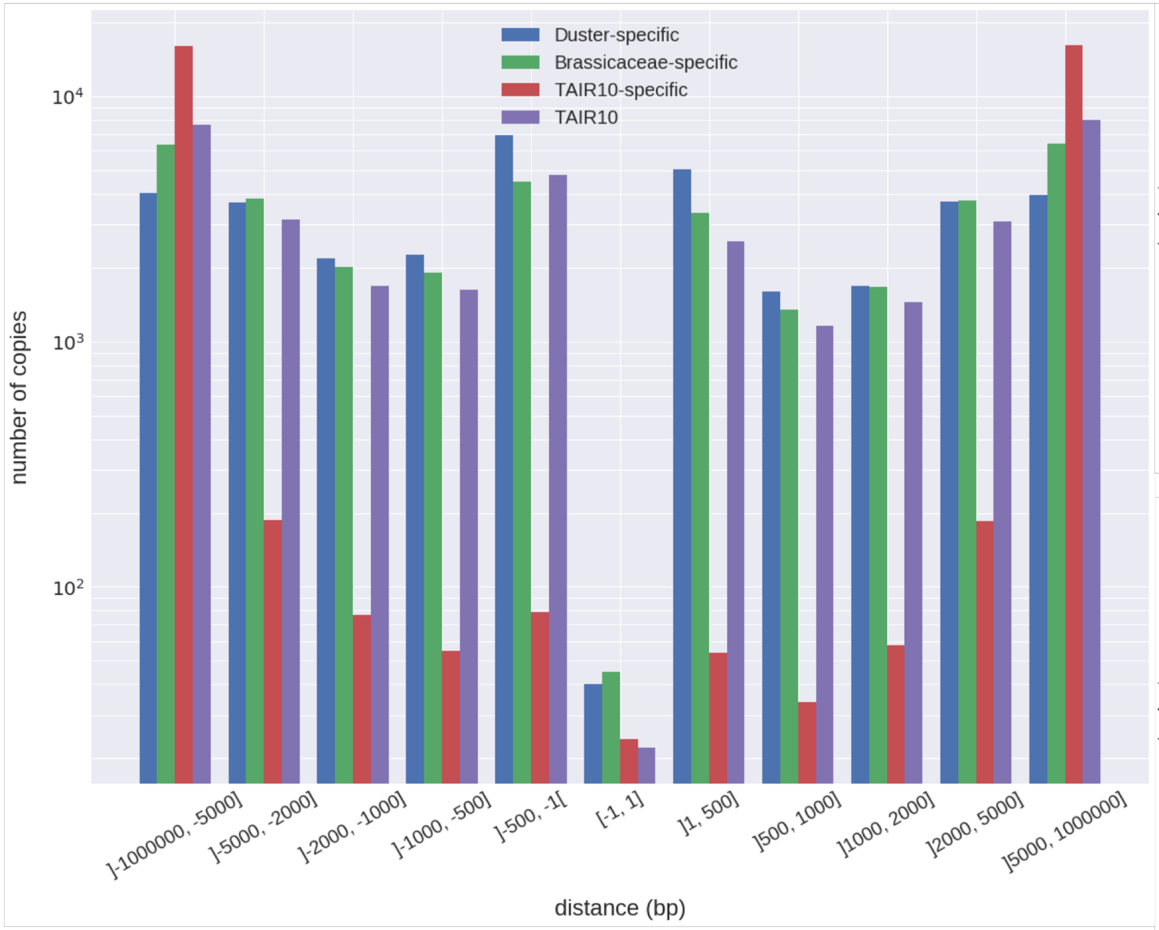


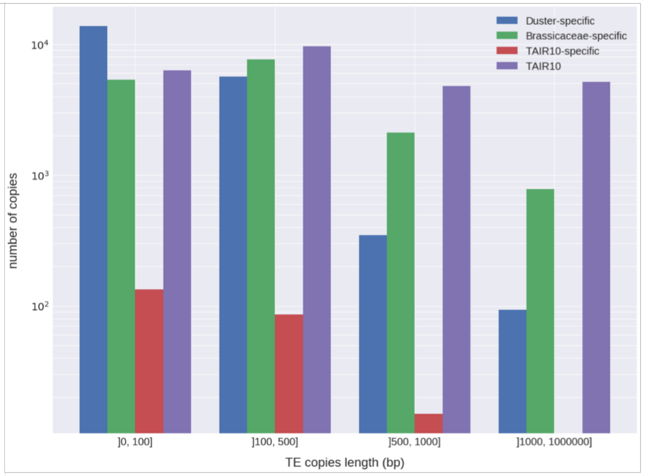
Combining annotations

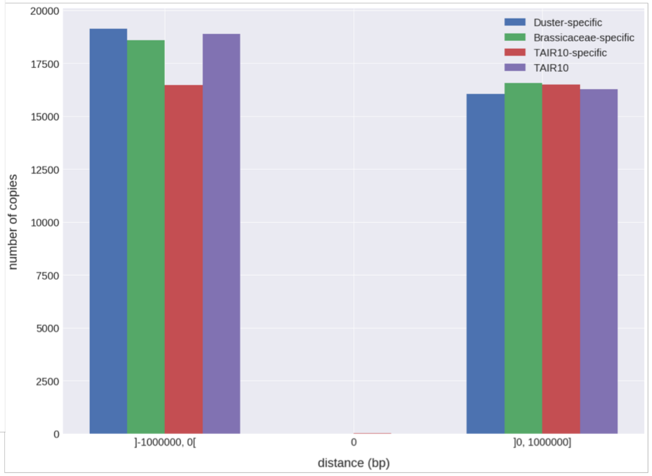


Hadi Quesneville

Location to closest gene

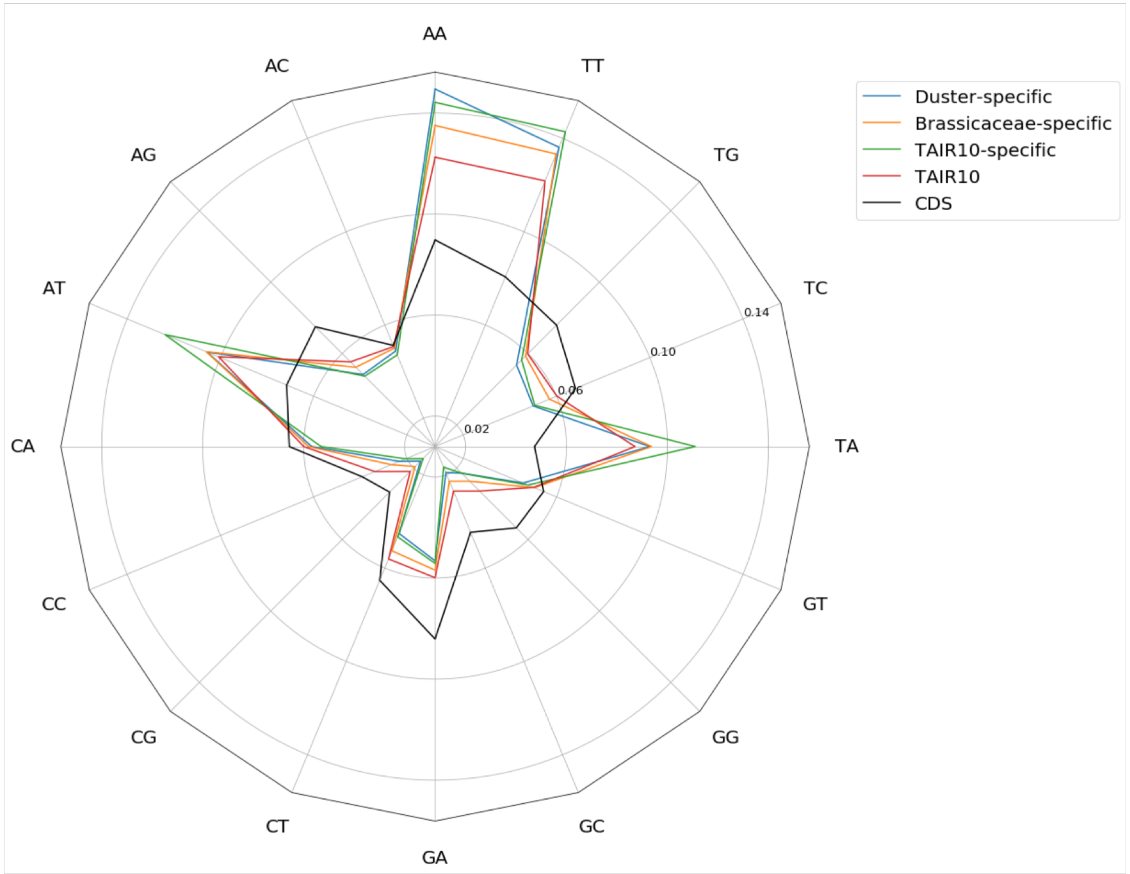






Hadi Quesneville

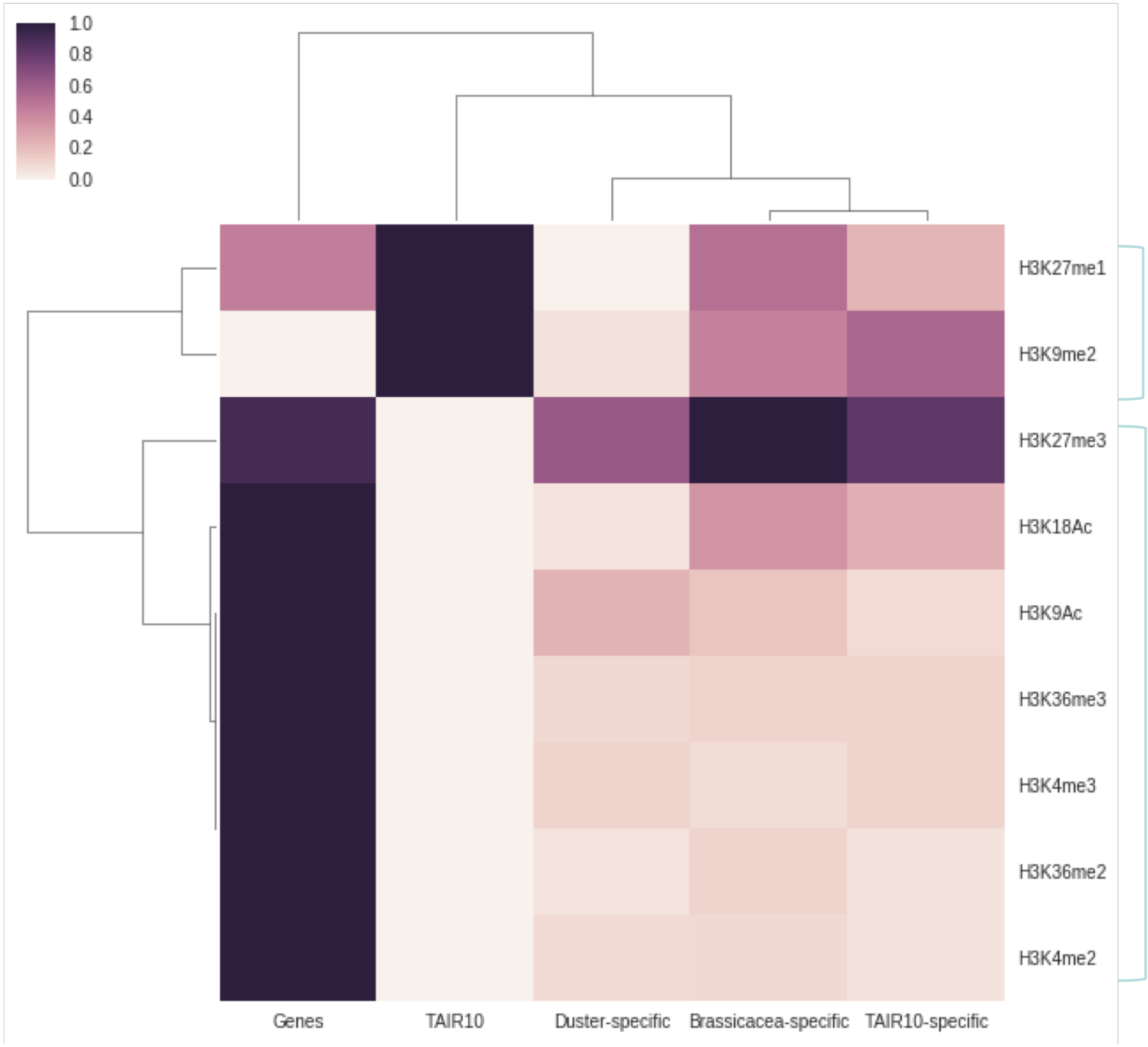
Base composition



Hadi Quesneville

Epigenetic status

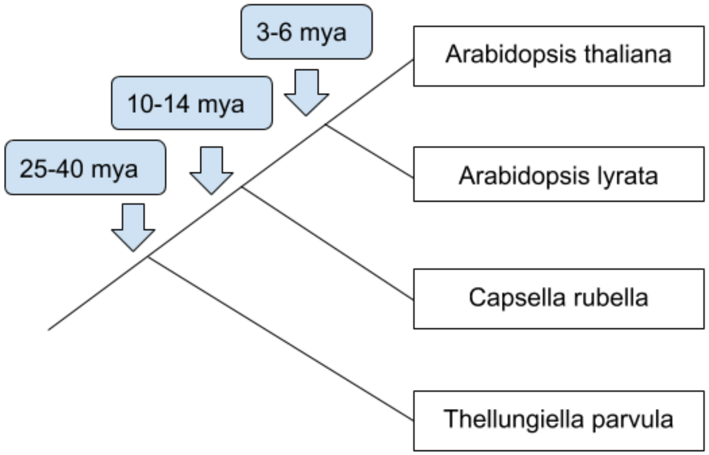
Luo *et al.*, 2009. Defining the Functional Network of Epigenetic Regulators in Arabidopsis thaliana. Molecular Plant 2, 661-674.

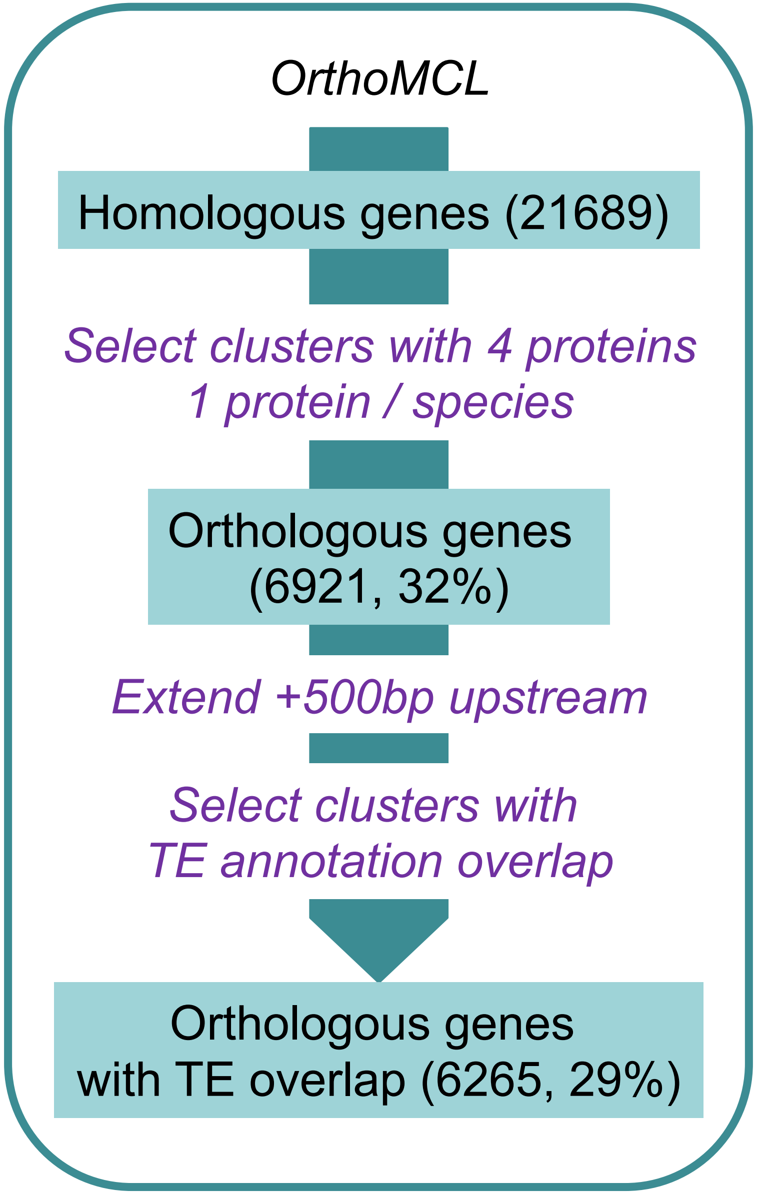


Heterochromatine

Euchromatine

Conservation → Orthologs





OrthoMCL

Homologous genes (21689)

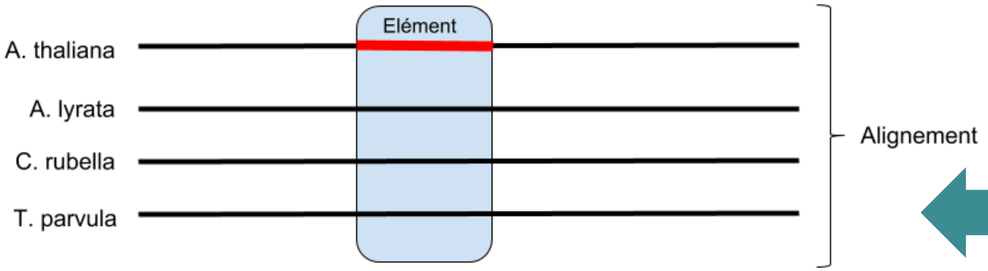
Select clusters with 4 proteins
1 protein / species

Orthologous genes
(6921, 32%)

Extend +500bp upstream

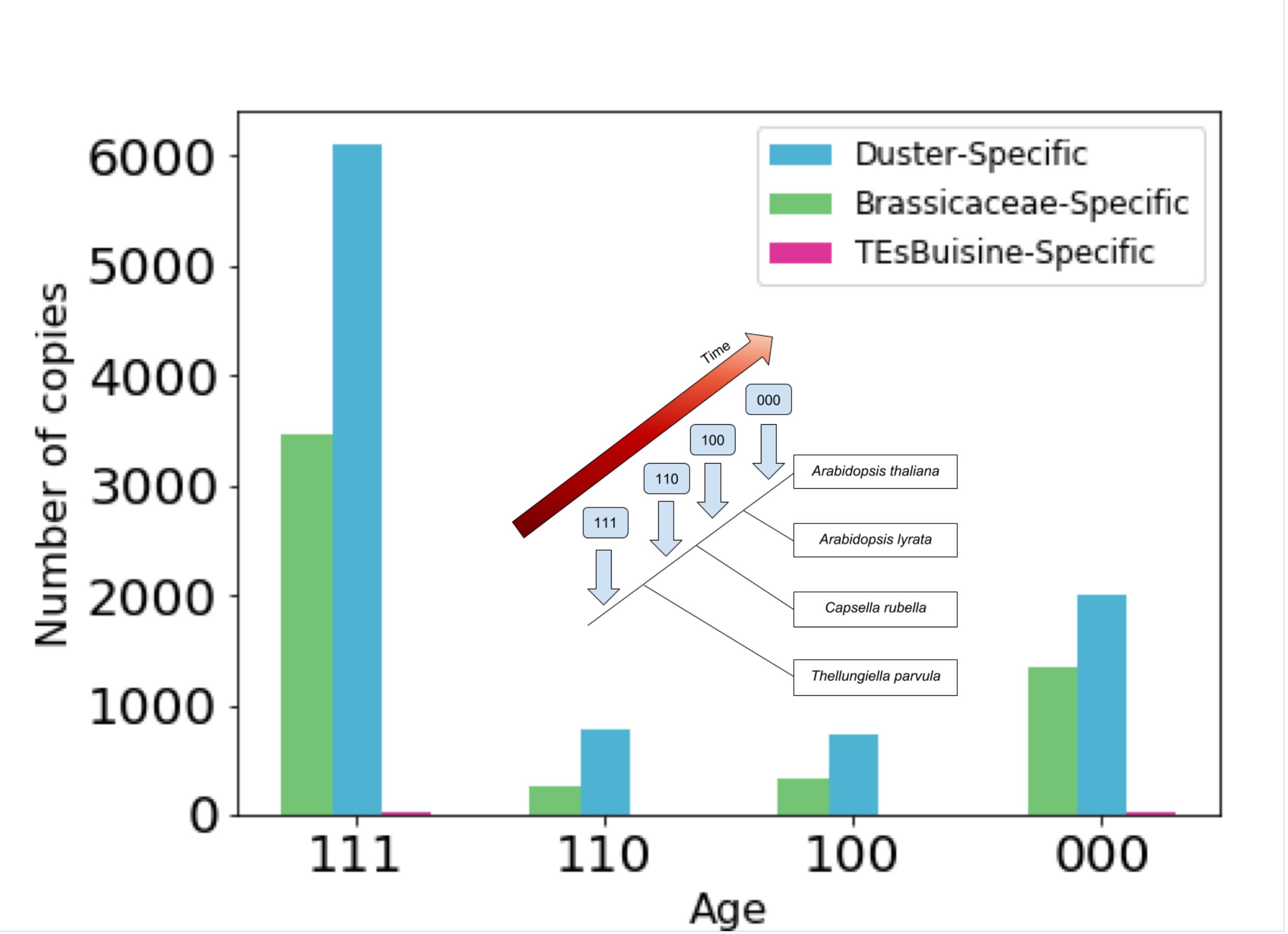
Select clusters with
TE annotation overlap

Orthologous genes
with TE overlap (6265, 29%)



50%?

Age of insertion



Hadi Quesneville

Conclusions

- ***Duster*** **more sensitive and faster than** ***BLAST*** **and** ***MegaBLAST (not shown in this presentation)***

- **Up to ~50% of** ***A. thaliana*** **genome derives from TEs.**
  - Official annotation ~20%

- **Found old and degenerated TE fragments**

- **Upstream of genes**

Thanks

The « Genome Analysis » team

Joëlle Amselem
Florian Maumus
Françoise Alfama
Nathalie Choisne
Johann Confais
Véronique Jamilloux
Hervé-Antoine Kerjean
Vikas Sharma
Agnès Baud
Mariène Wan

INSTITUT JACQUES MONOD
Danielle Nouaud
Dominique Anxolabéhère