Evolution of recombination landscape in diverging populations of bread wheat

Danguy des Déserts A. , Bouchet S., Sourdille P., Servin B.
Genome Biology Evolution, 2021

UNIVERSITÉ Clermont Auvergne

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

GDEC
Génétique Diversité Ecophysiologie des Céréales

GenPhySE
Génétique, Physiologie et Systèmes d'Elevage

FLORIMOND DESPREZ

Wheat production faces many challenges

IWGSC **Fact Sheet**

www.wheatgenome.org

Wheat is the most widely grown cereal crop in the world and the staple food for more than 35% of the human population

730 million tons produced

$50 billion of total export value

222 million hectares harvested

60% increase in production needed to feed 9.6 billion by 2050

Wheat FACTS

486 million tons used for human consumption

2.5 billion consumers in **89** countries

19% of our total available calories

67 kilograms consumed on average per person per year

Data year: 2014

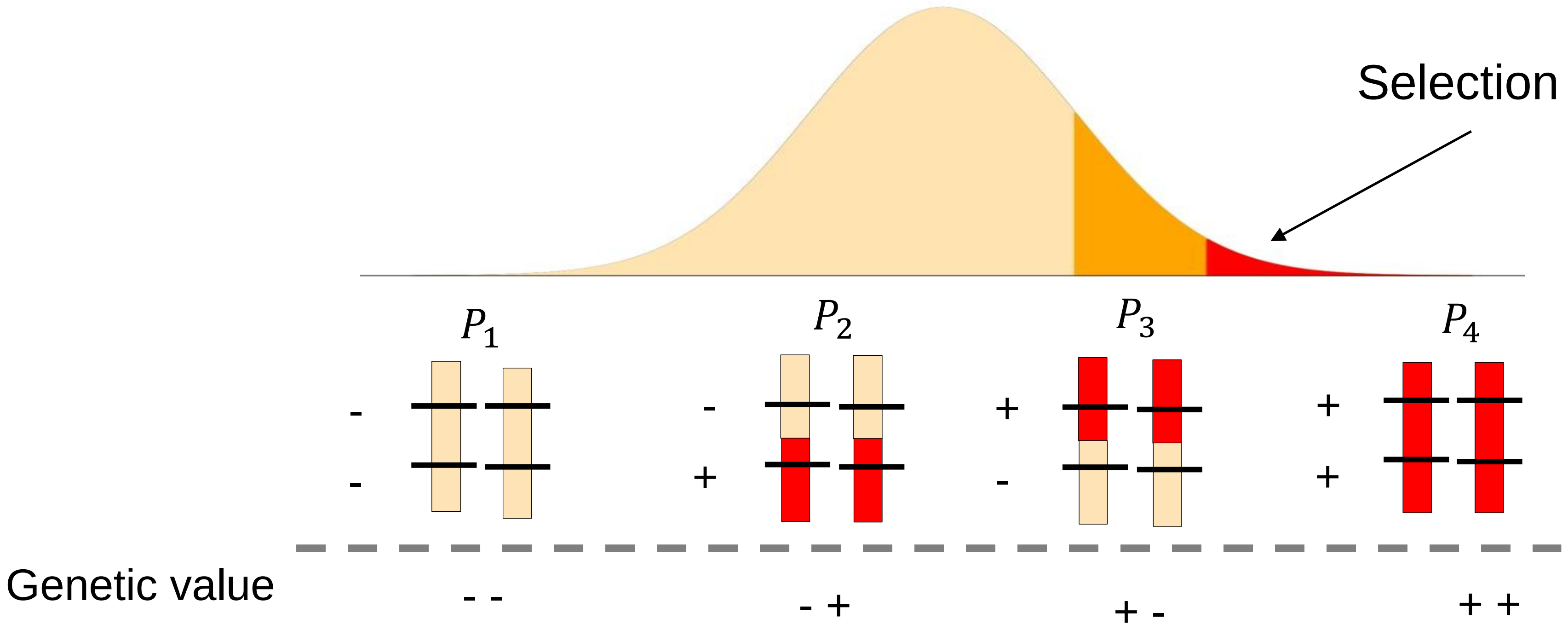
Sources: FAOSTAT www.faostat.fao.org ; CIMMYT www.wheat.org

Updated October 2016

- Demand for yield is increasing
- Reduction of intrants
- Emerging stresses

→ Varietal improvement

Varietal improvement relies on genetic variance

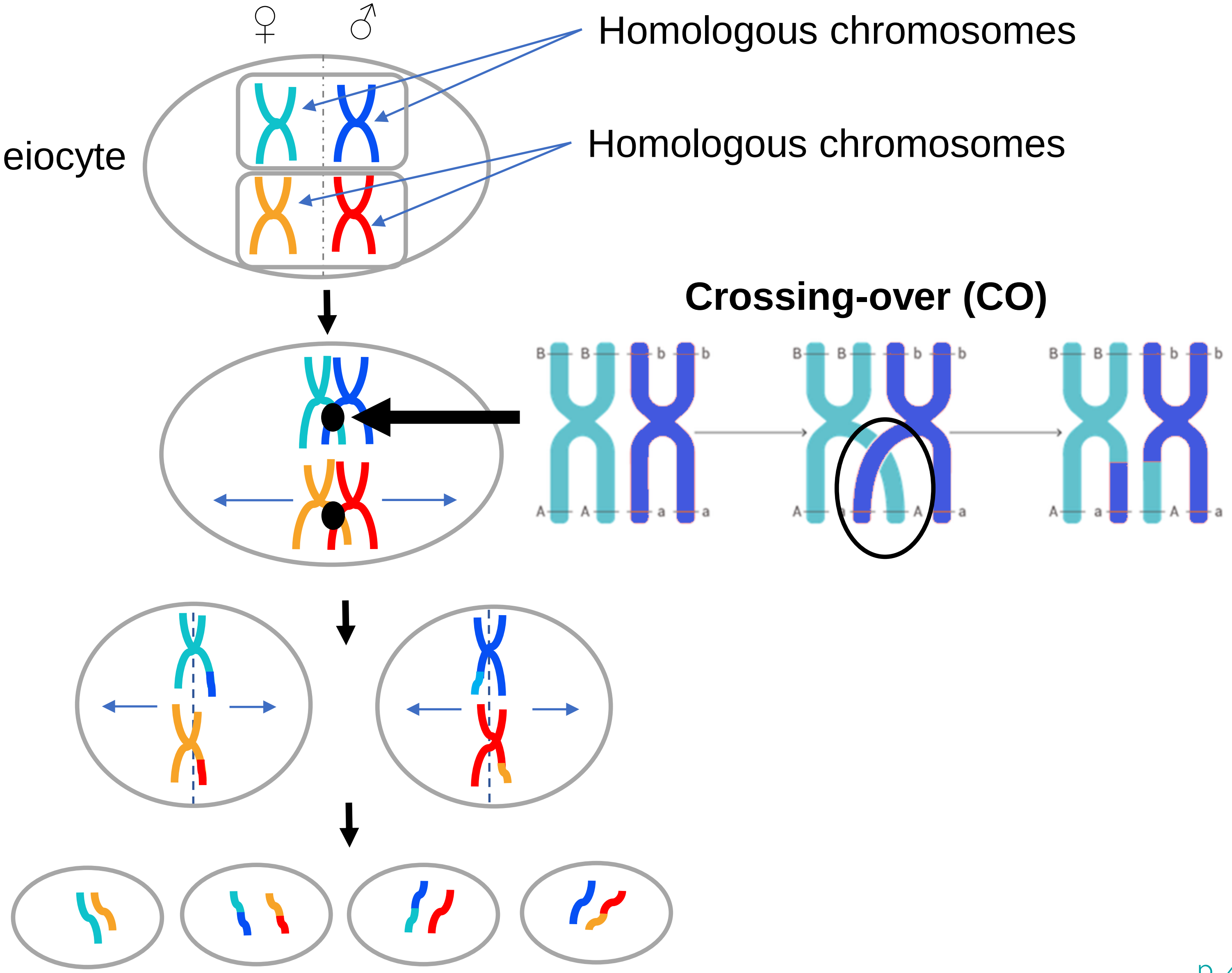


Genetic diversity results from

- Mutation
- Reshuffling of alleles through meiotic recombination

Meiotic recombination

Meiosis = production of gametes



meiocyte

Homologous chromosomes

Homologous chromosomes

1) Metaphase I

Crossing-over (CO)

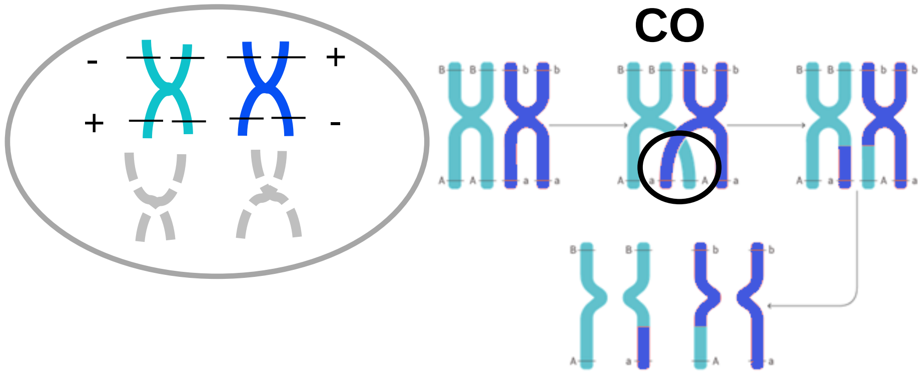
2) First division

3) Second division

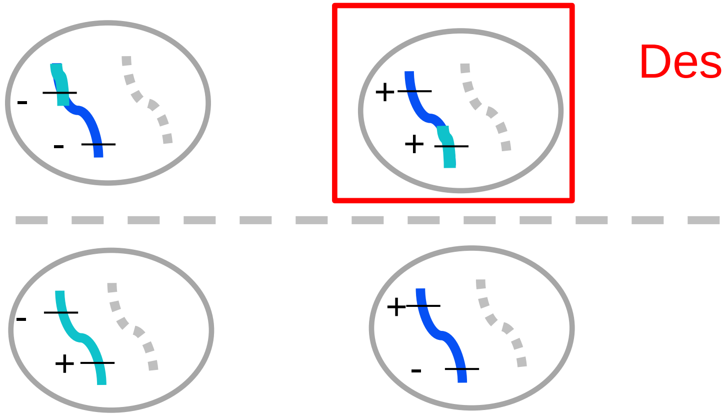
4 gametes

Recombination allows to cumulate desirable alleles in new varieties

If genes are on the same chromosome



Recombined
$f = r$



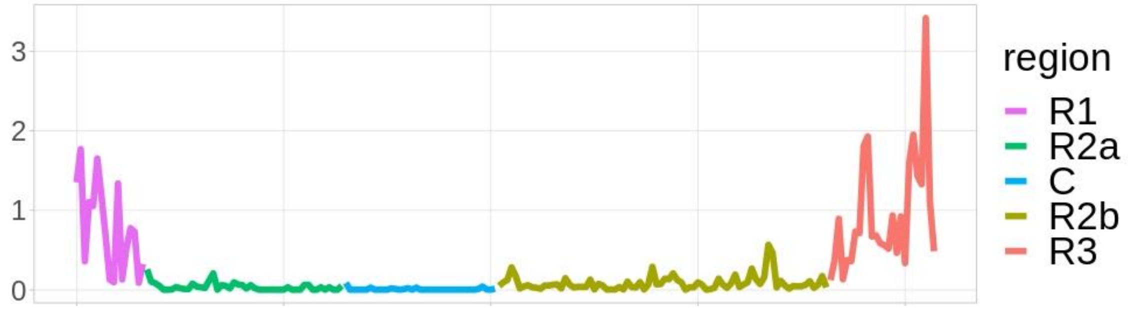
Desirable gamete
$p = r/2$

Not recombined
$f = 1 - r$

Recombination landscape (= repartition of CO along the genome)

Variability along the genome?

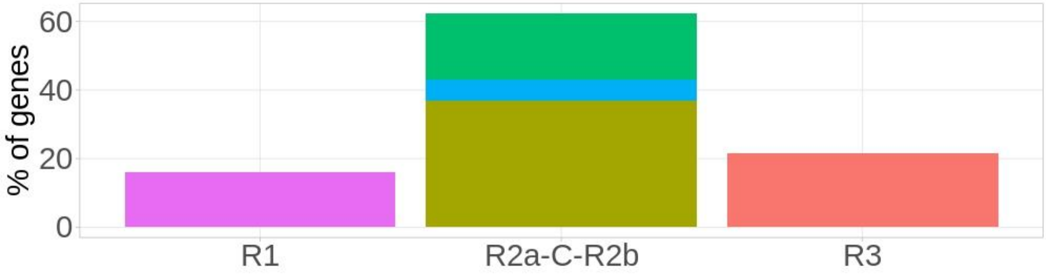
Recombination landscape (bread wheat)



chromosome 3B

Few CO in gene-rich regions
→ CO can be a limiting factor of varietal improvement

Genes count



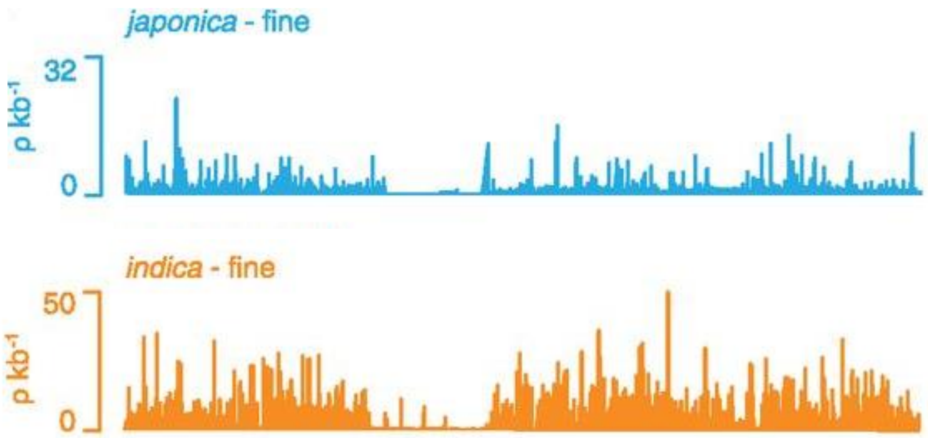
What drives recombination landscape ?

Recombination landscape

Recombination landscape = many mecanisms (interference, DNA compaction…)

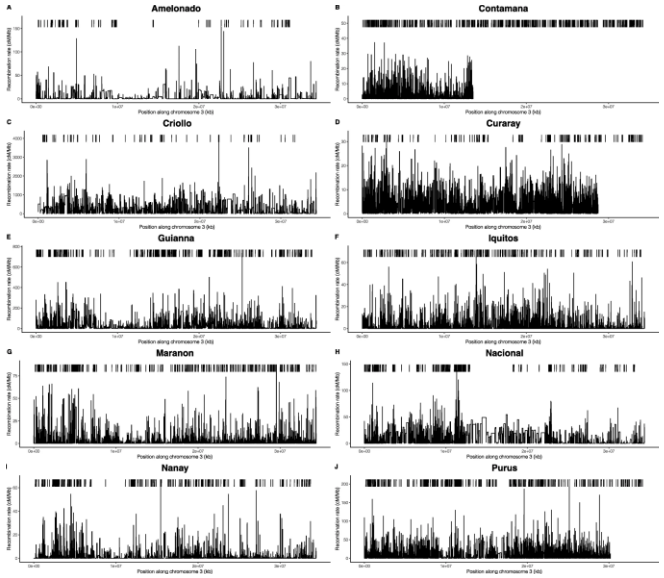
One approach to caracterize the determinism of rec. landscape = variability between populations

Marand et al. 2019 (The Plant Cell):
Comparison of landscapes of two **rice** sub-species



80% of recombination hotspots are **<u>sub-species specific</u>**

Schwarzkopf et al. 2020 (BMC genomics):
Comparison of landscapes of ten **cacao-tree** populations



55% of hotspots are **<u>population specific</u>**

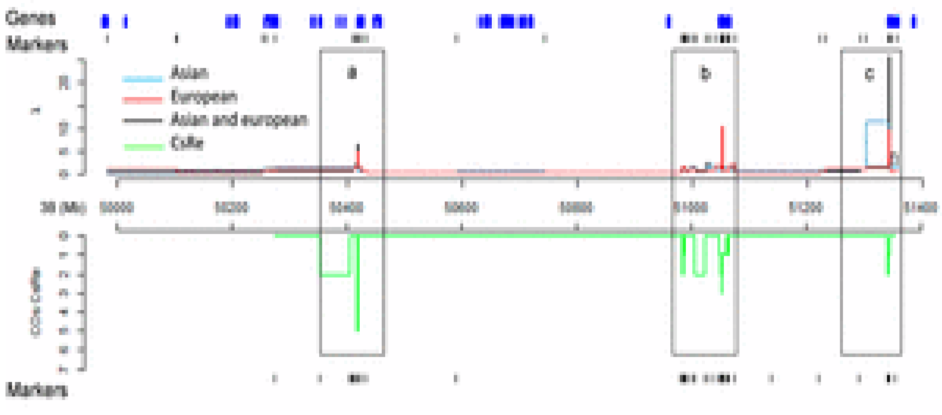
Variability of recombination landscape between populations ?

In bread wheat ?

Gardiner et al. 2019 (Genome Biology):

- Comparison of recombination landscape in 13 NAM families (1 parent x 13 other parents)
- More similar parents have more similar recombination landscapes

Darrier et al. 2017 (Genetics):



🟦 Asian population
🟥 European
⬛ Asian + European
🟩 CsRe

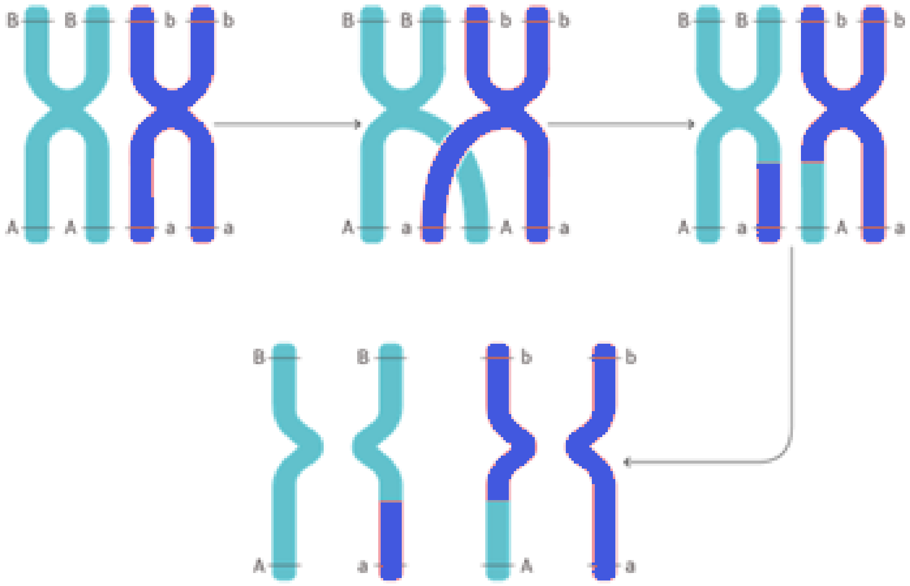
- Recombination landscapes are ~ similar
- But analysed region < 4 Mb

This presentation:
Genome-wide variability of recombination landscape between diverging populations of bread wheat

Methods

Estimate recombination landscape

Method 1: Count recombination within families: « Meiotic landscape »
Ex: Chinese Spring * Renan population (Rimbert et al. 2018, PloS one)



Drawbacks:

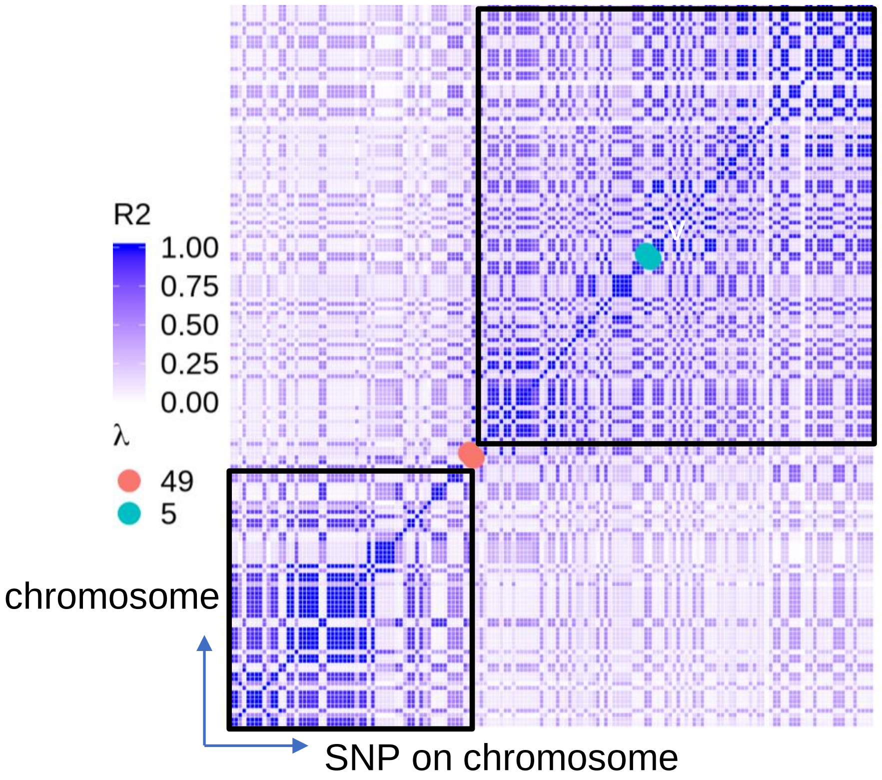
- **Small number of progeny**
- **Specific of parents**

Recombination rate $\hat{c} = f\left(\frac{nb\ recombined\ gametes}{nb\ gametes}\right)$

Estimate recombination landscape

Method 2: linkage disequilibrium of diversity panel: « LD-based landscape »

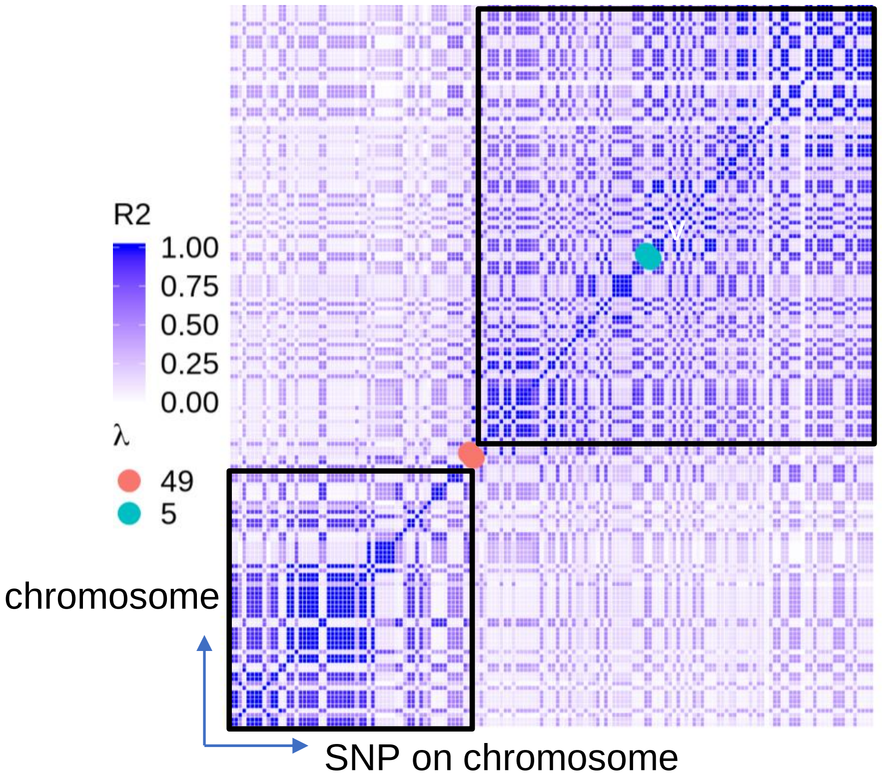
LD patterns



Estimate recombination landscape

Method 2: linkage disequilibrium of diversity panels: « LD-based landscape »

LD patterns

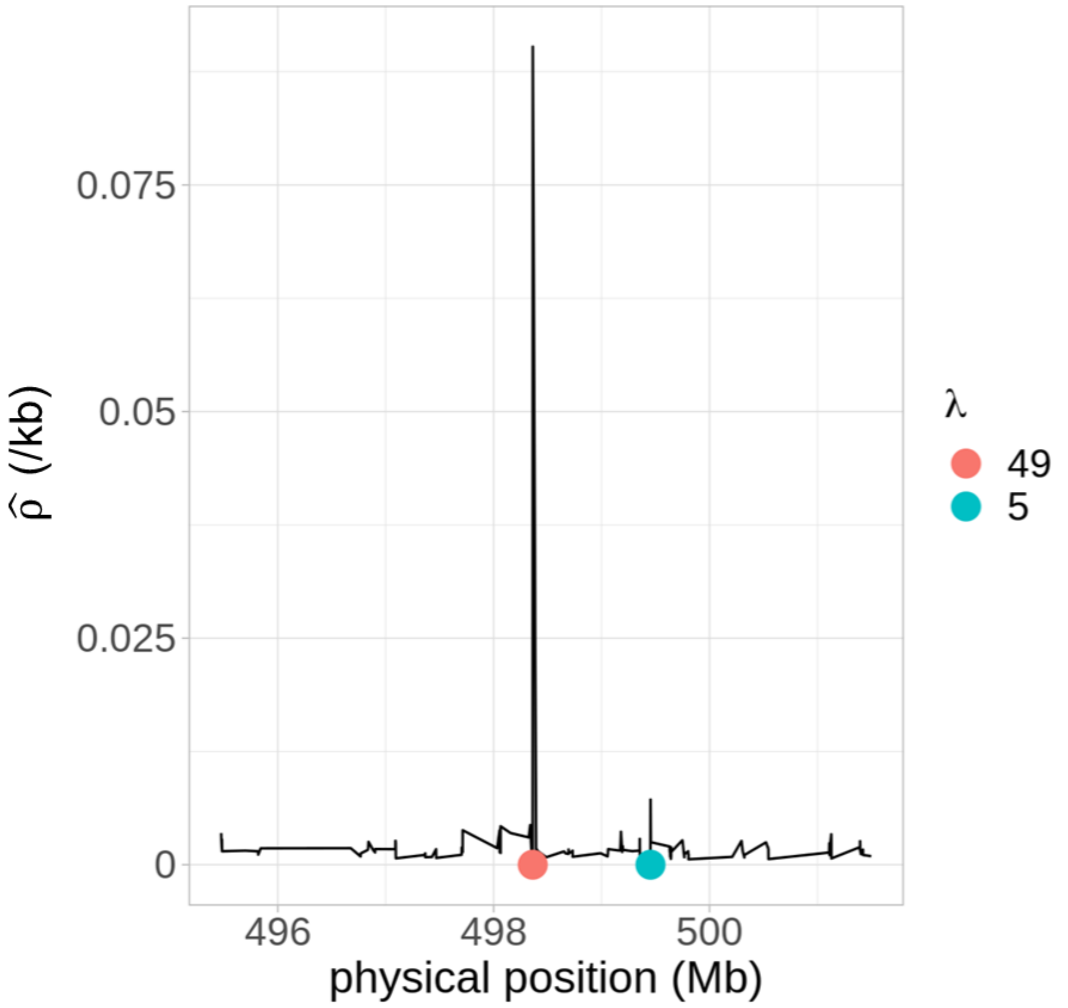


Li et Stephens model (2003, Genetics)

$$\hat{\rho} \approx \rho = K * c$$

with K = f(Ne)

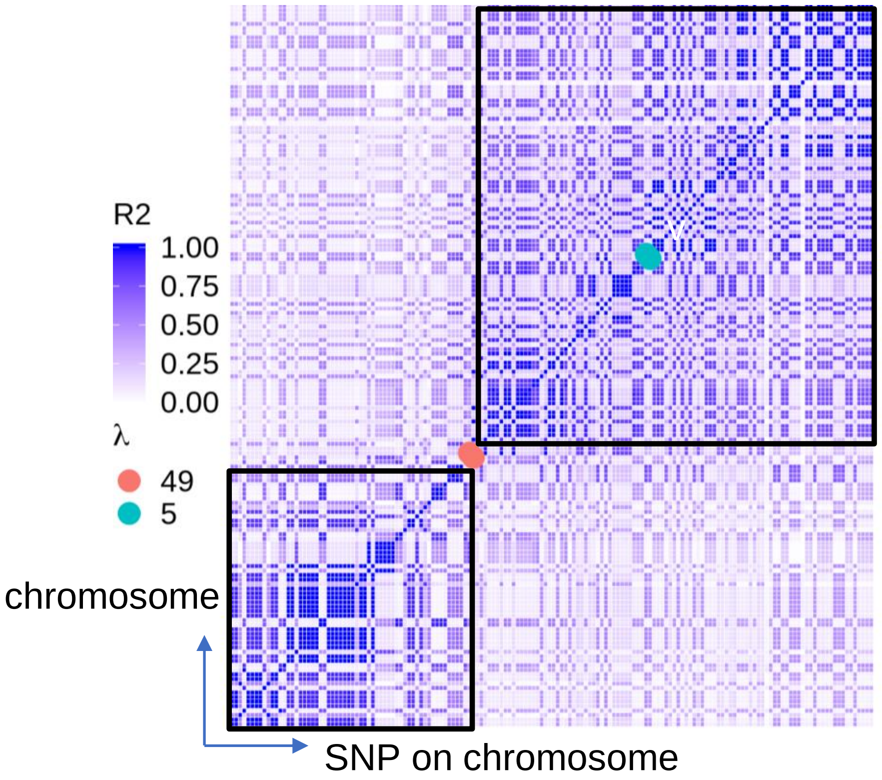
LD-based recombination landscape



Estimate recombination landscape

Method 2: linkage disequilibrium of diversity panels: « LD-based landscape »

LD patterns



Li et Stephens model (2003)

$$\hat{\rho} \approx \rho = K * c$$

with $K = f(Ne)$

Advantages compared to meiotic method:

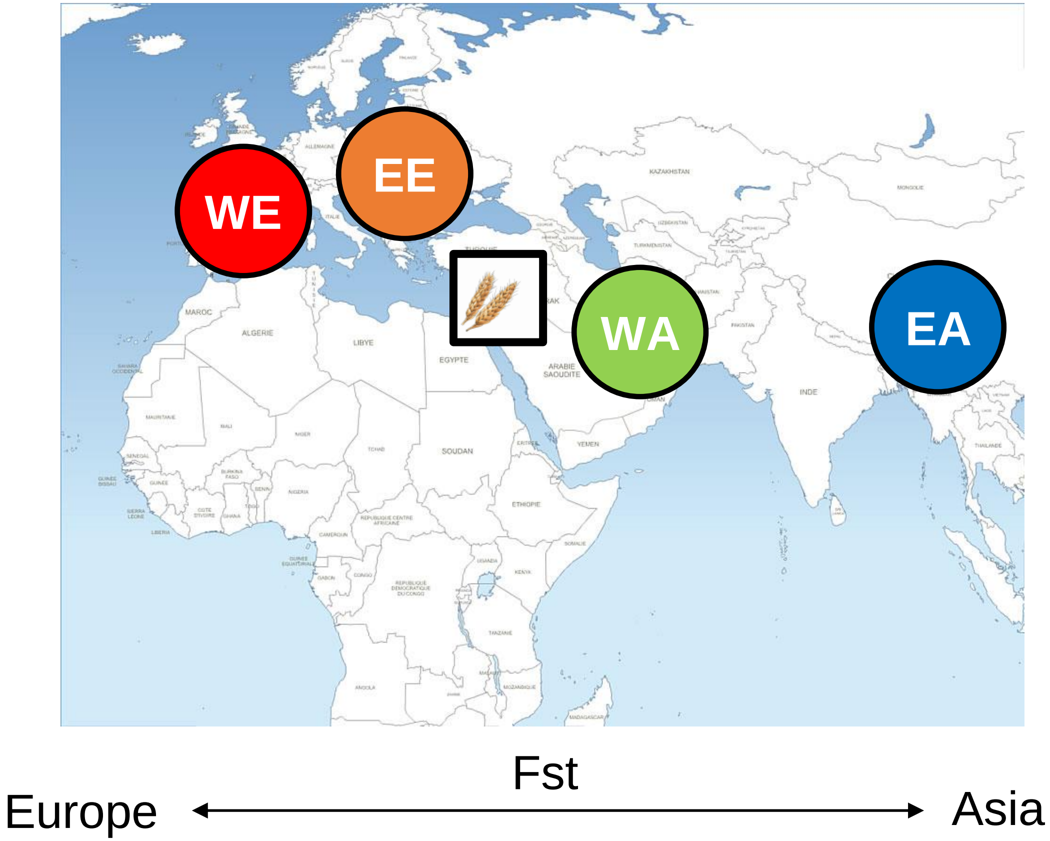
- **Based on many meiosis**
- **Many polymorphic SNP**
- **Representative of the population**

→ Better suited to compare recombination landscapes between populations

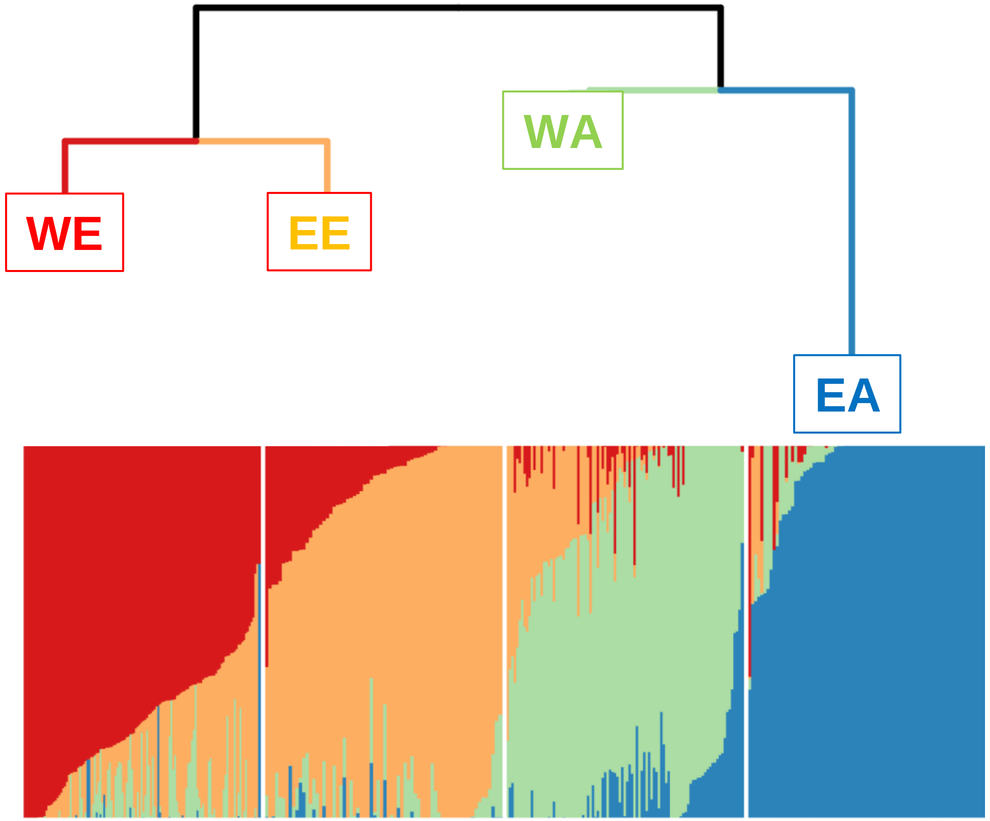
4 different populations of bread wheat

371 bread wheat landraces sampled worldwide (Balfourier et al. 2019, Science Advances)
130k SNP of TABW410k (Kitt et al. 2021, Zenodo)
→ hierarchical clustering K = 4

Populations



Differenciation of populations



- ➢ Do the recombination profiles of these 4 populations vary ?

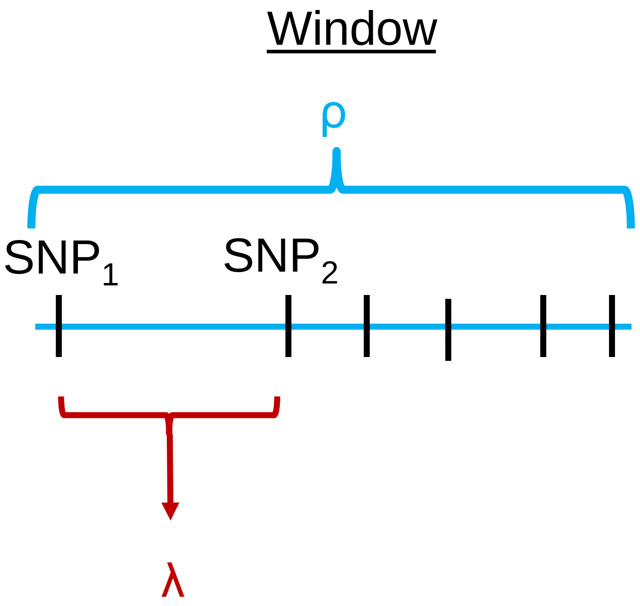
Estimate recombination landscapes

For each population (WE, EE, WA, EA):

- Split the genome into ~ 600 windows of ~ 2 cM length (using CsRe)

- Run PHASE (Li et Stephens 2003, Genetics)

= to obtain joint posterior distributions of ρ and λ



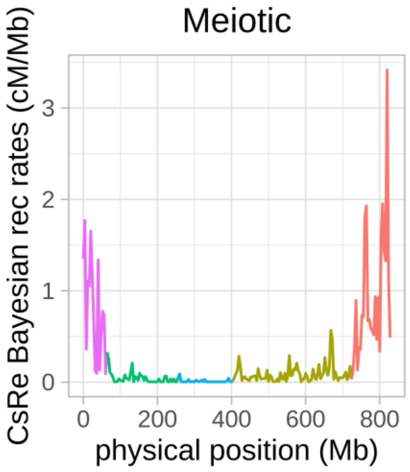
ρ : Background recombination rate

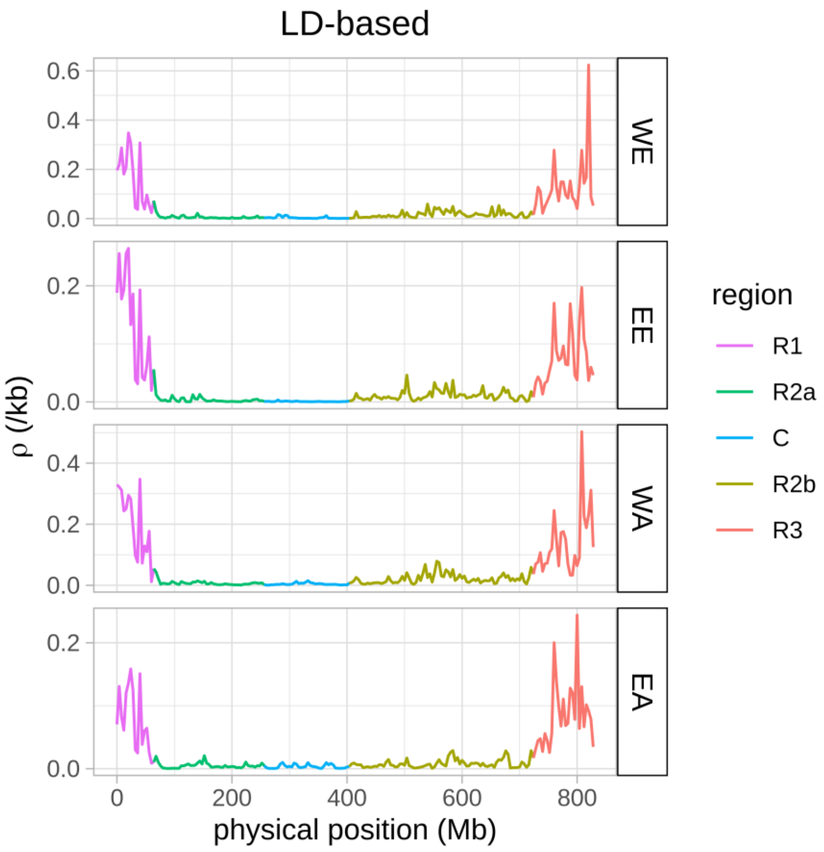
λ : Local inflation of recombination

Results

Validation of LD-based recombination landscapes

3B chromosome



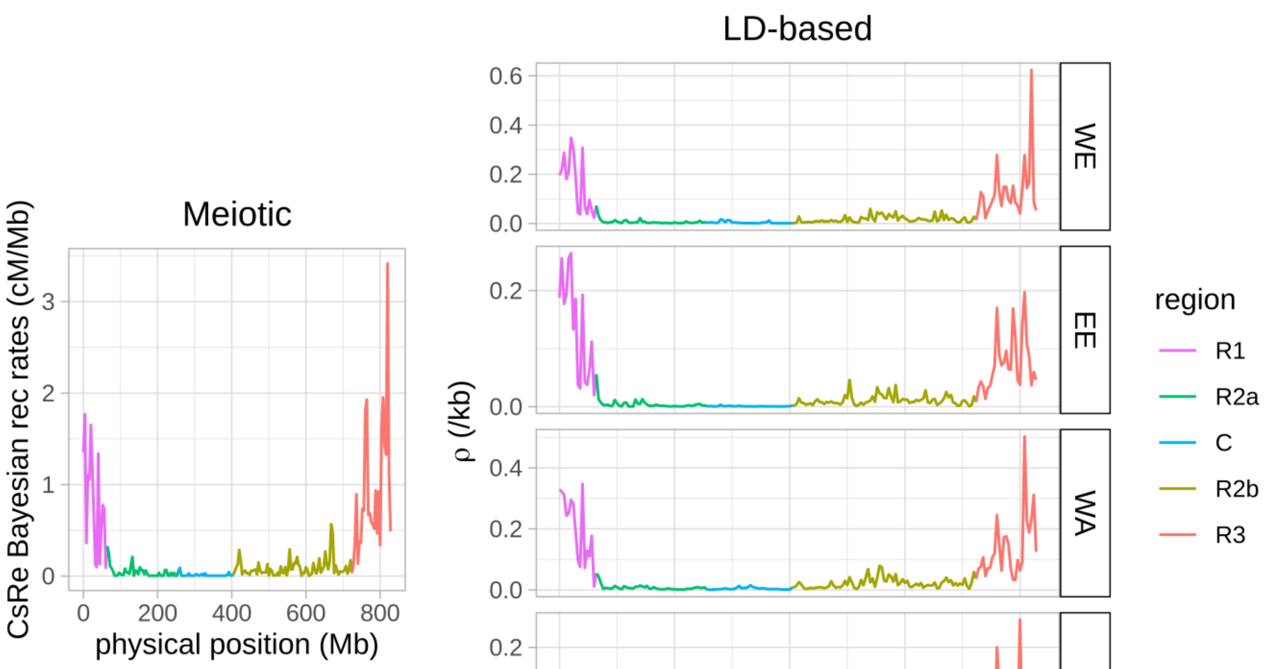


$$\hat{c} = f\left(\frac{nb\ recombined\ progenies}{nb\ progenies}\right)$$

$$\hat{\rho} = K*c$$

Validation of LD-based recombination landscapes

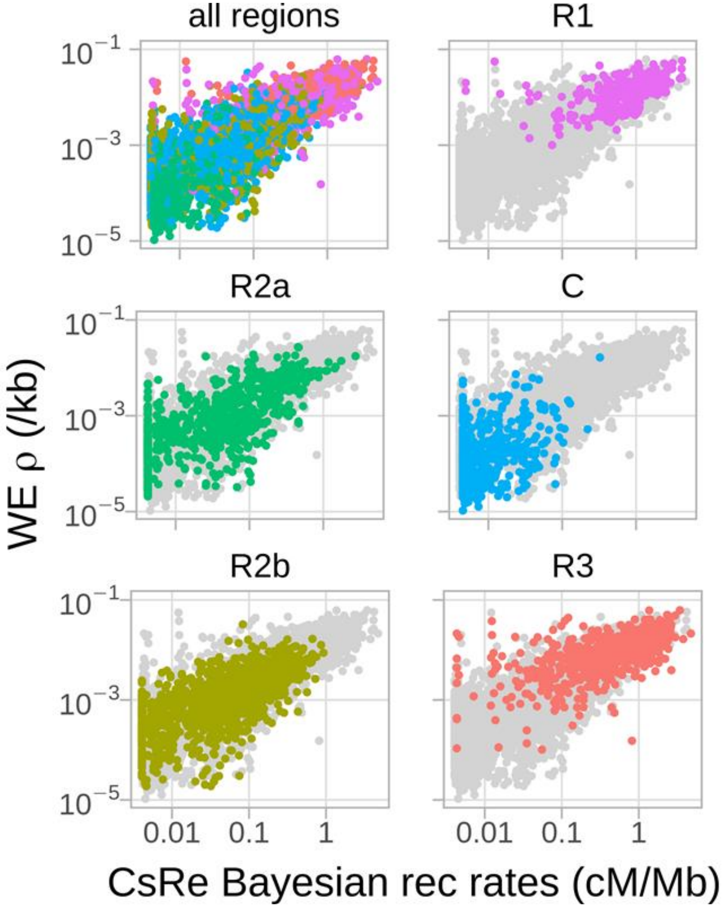
3B chromosome



$$\hat{c} = f\left(\frac{nb\ recombined\ progenies}{nb\ progenies}\right)$$

$$\hat{\rho} = K*c$$

Correlation all genome

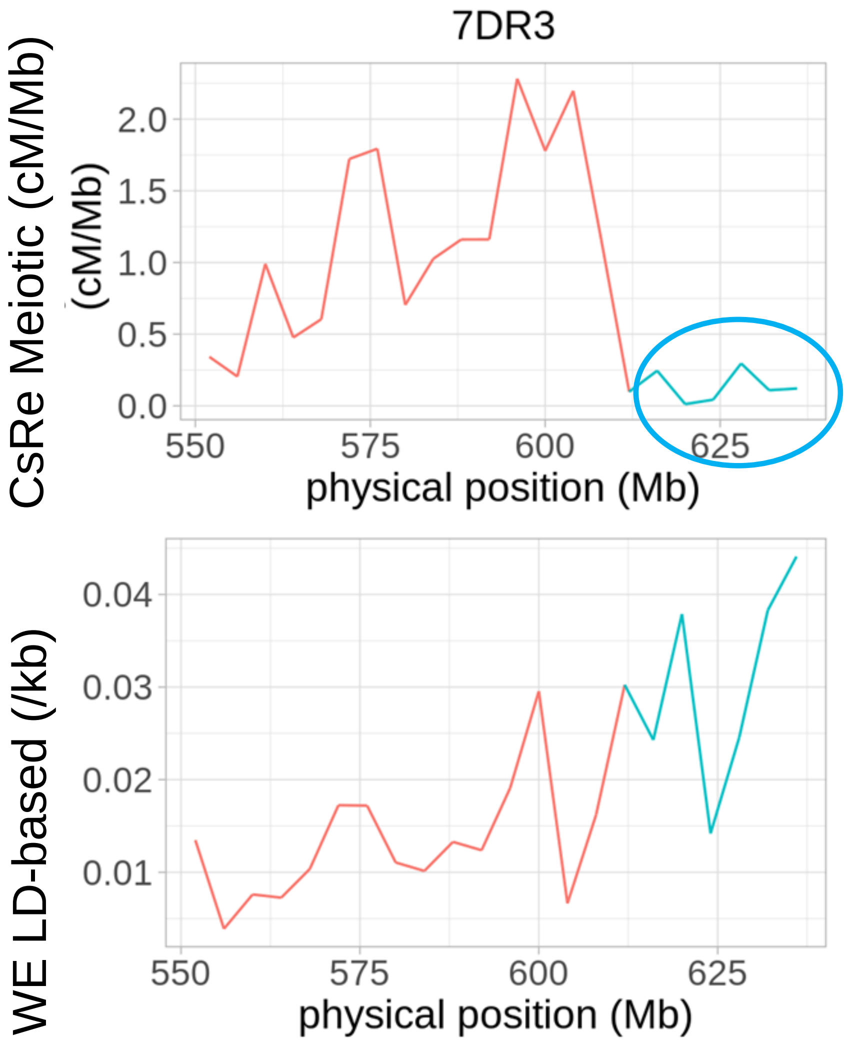


Average correlation within genomic regions (1AR1…7DR3) = 0.6 ± 0.2

Local differences between meiotic and LD-based landscapes

- Low correlation in the 7DR3 region
- Wild introgression in Renan population

→ Meiotic landscapes are sensitiv to individual specific variation

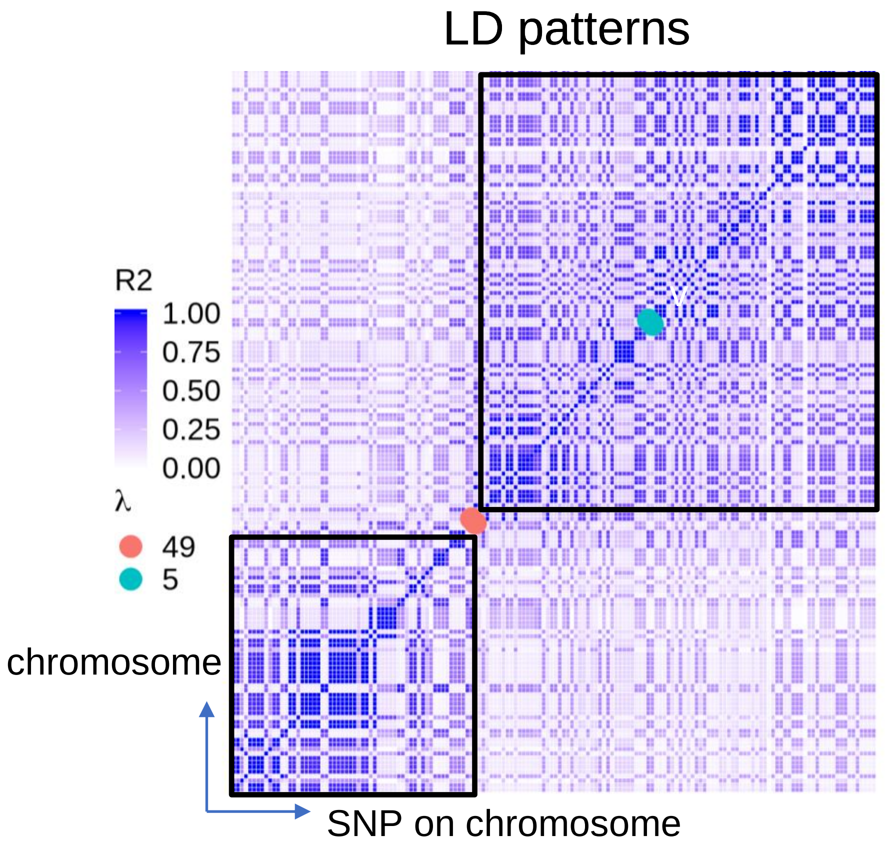


Validation of LD-based recombination landscapes

- Good correlation between LD-based and meiotic landscapes
- But higher resolution in LD-based landscapes

→ LD-based = better suited to study and compare
fine scale variation of recombination between the 4 populations

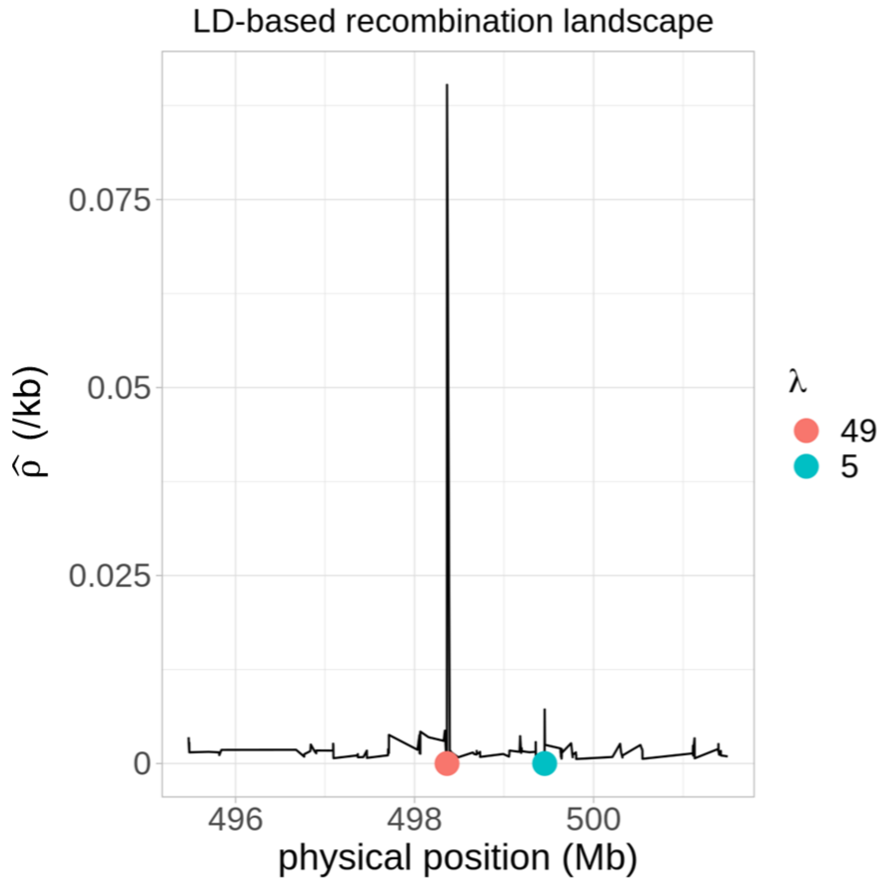
Comparison of fine scale variation of LD-based landscapes



Li et Stephens model (2003)

$\hat{\rho} \approx \rho = K * c$

$\xrightarrow{\text{with } K = f(Ne)}$

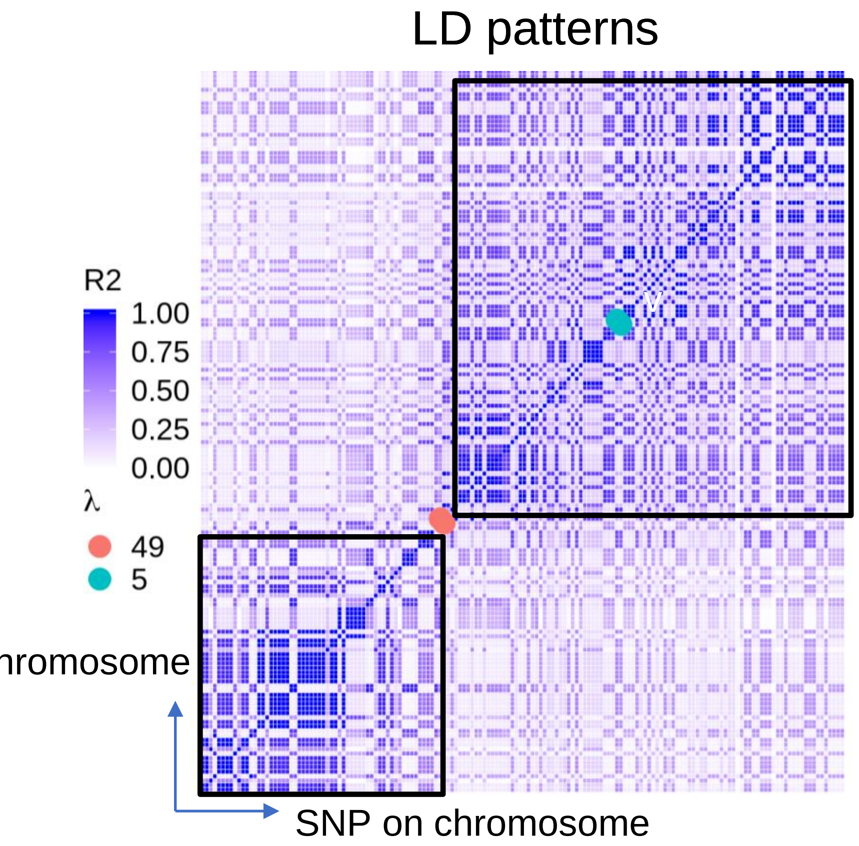


Drawback: $\hat{\rho}$ sensitivity to evolutionary forces K= f(Ne)

Example: selection acts locally

$$\lambda = \frac{\rho_{interval}}{\rho_{larger\ window}} = \frac{K * c_{interval}}{K * c_{window}}$$ = local inflation of recombination

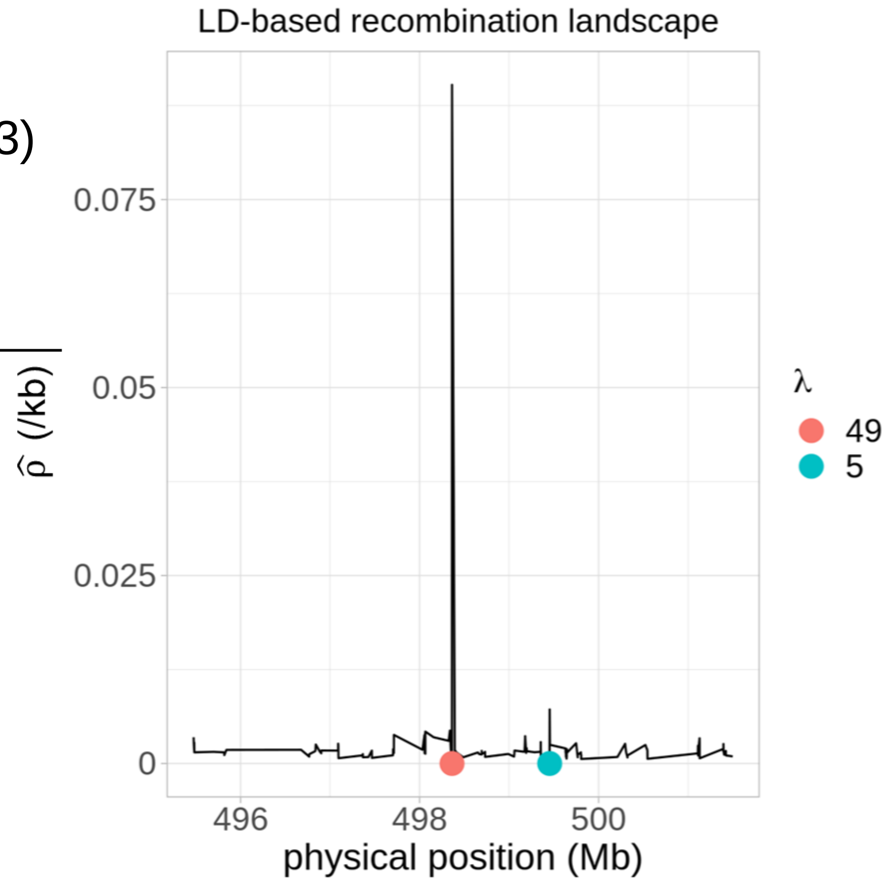
Fine scale variation of LD-based landscapes



Li et Stephens model (2003)

$\hat{\rho} \approx \rho = K * c$

with K = f(Ne)



$$\lambda = \frac{\rho_{interval}}{\rho_{larger\ window}} = \frac{K * c_{interval}}{K * c_{window}}$$ = local inflation of recombination

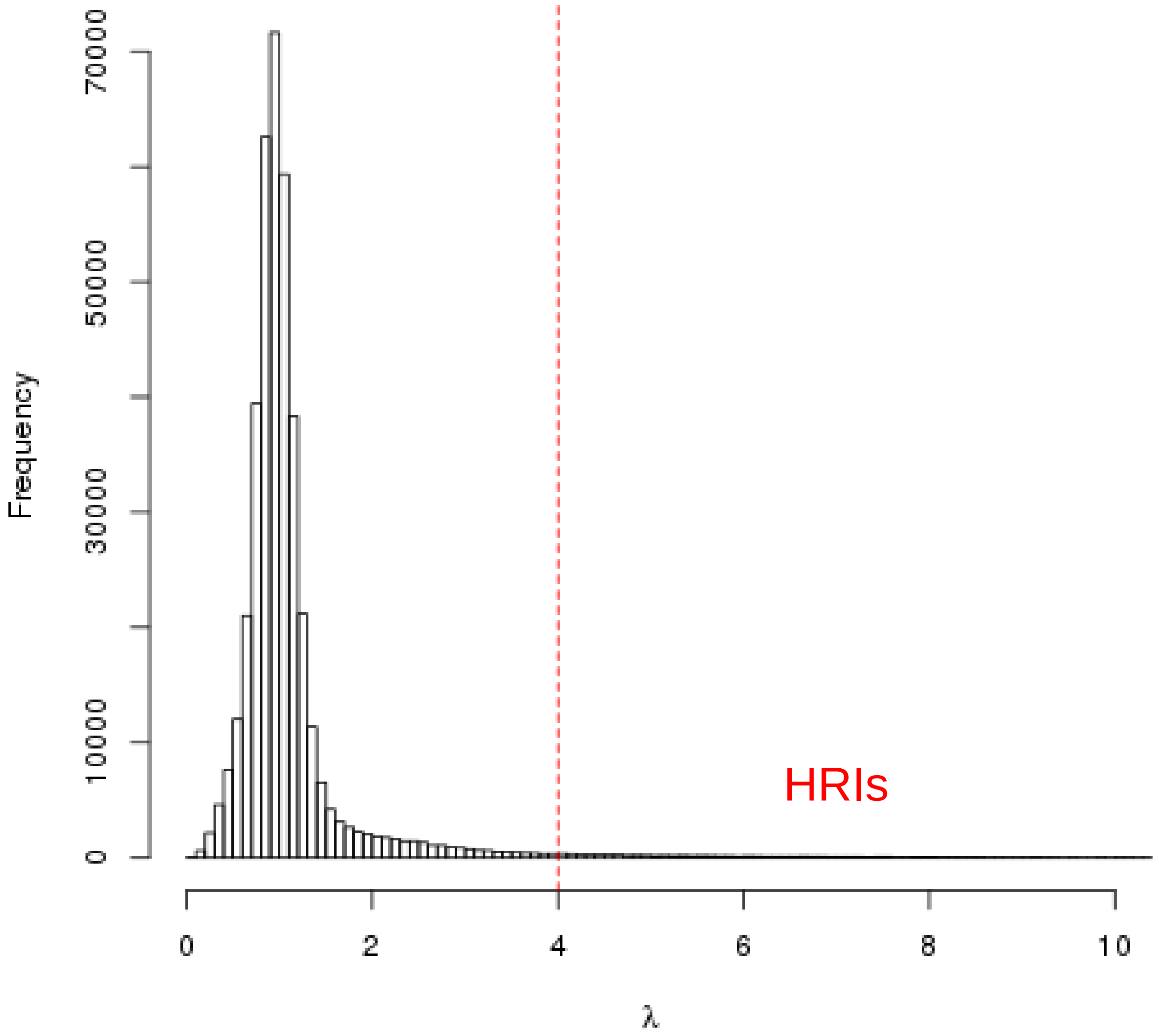
λ landscape = variation of recombination « only »
→ λ = better suited to compare populations

Localization of high recombination rates

Highly Recombining Intervals (HRIs)

HRIs = intervals with high λ

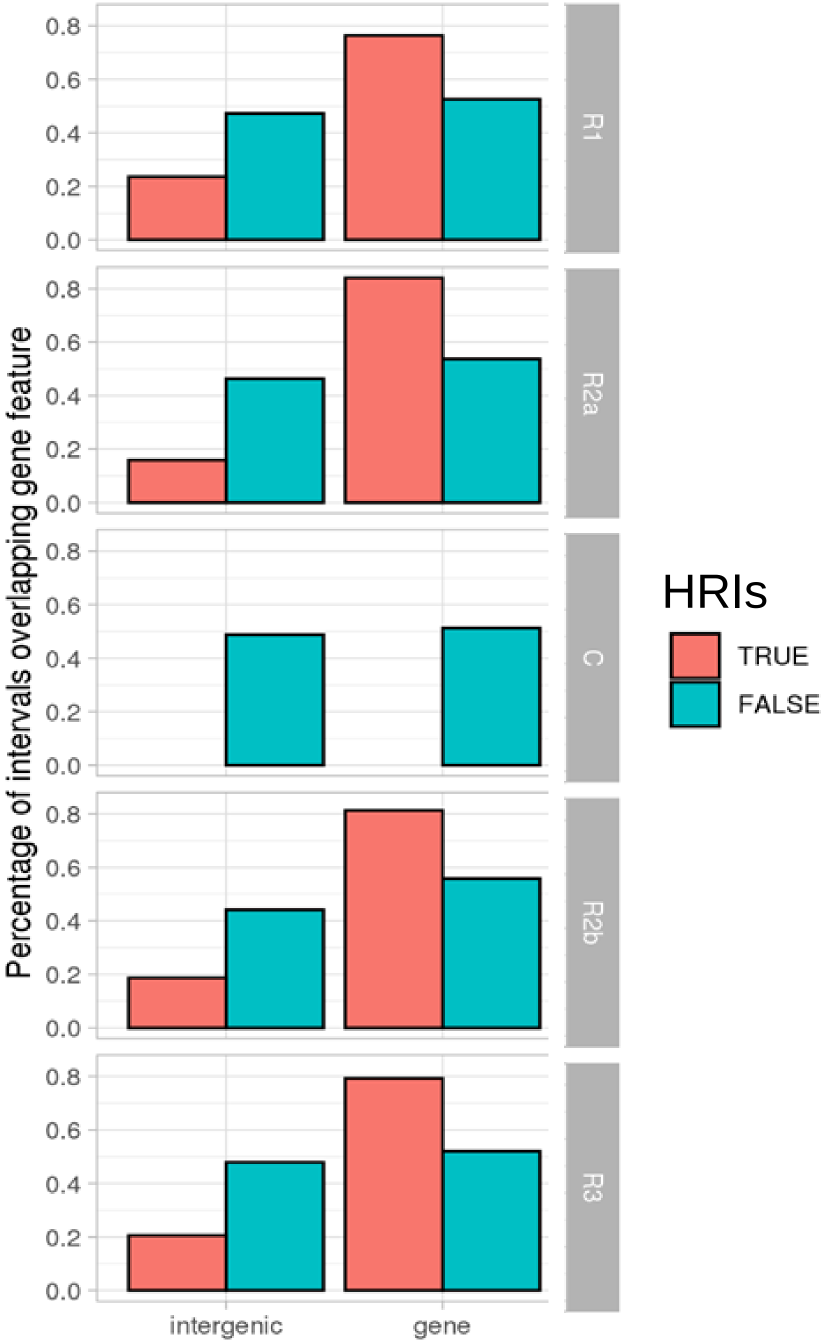
$\lambda = \frac{c_{interval}}{c_{window}}$ = local inflation of recombination



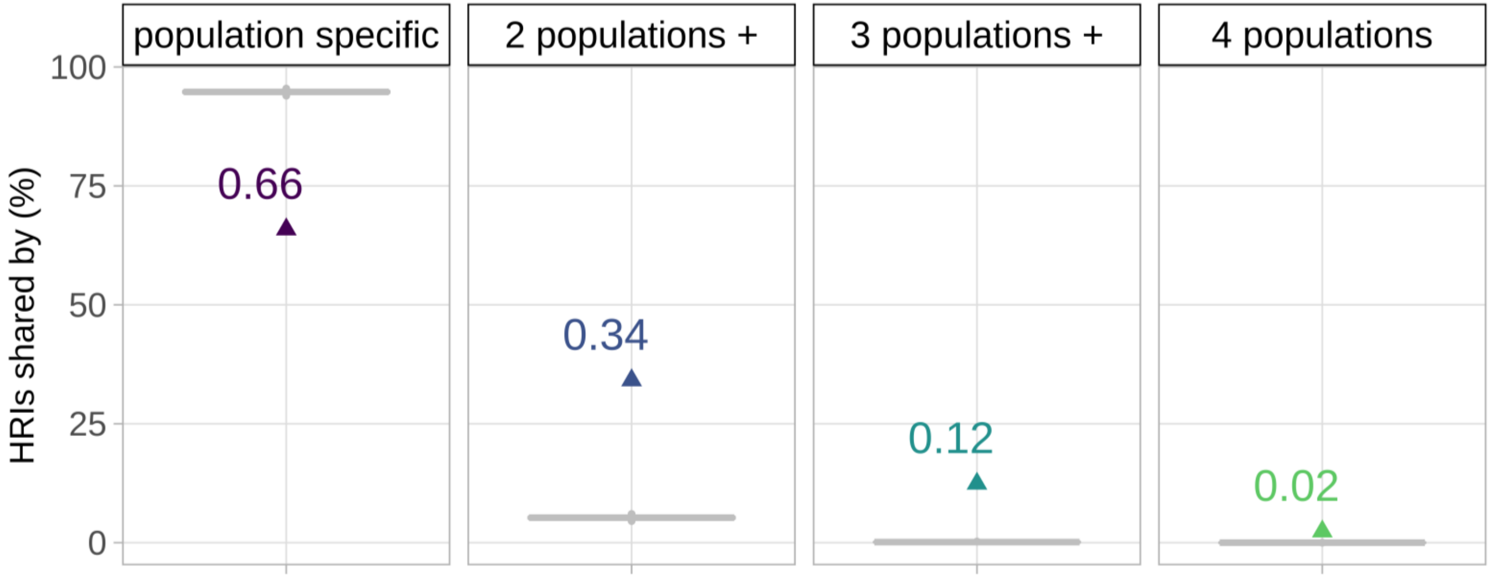
Features associated with HRIs

- ~ 9k HRIs (all populations combined)
- > 70% of HRIs located in regions R1 and R3
- 80% of HRIs overlapped genes

→ HRIs associated with open-chromatine
Consistent with litterature on rec. hotspots



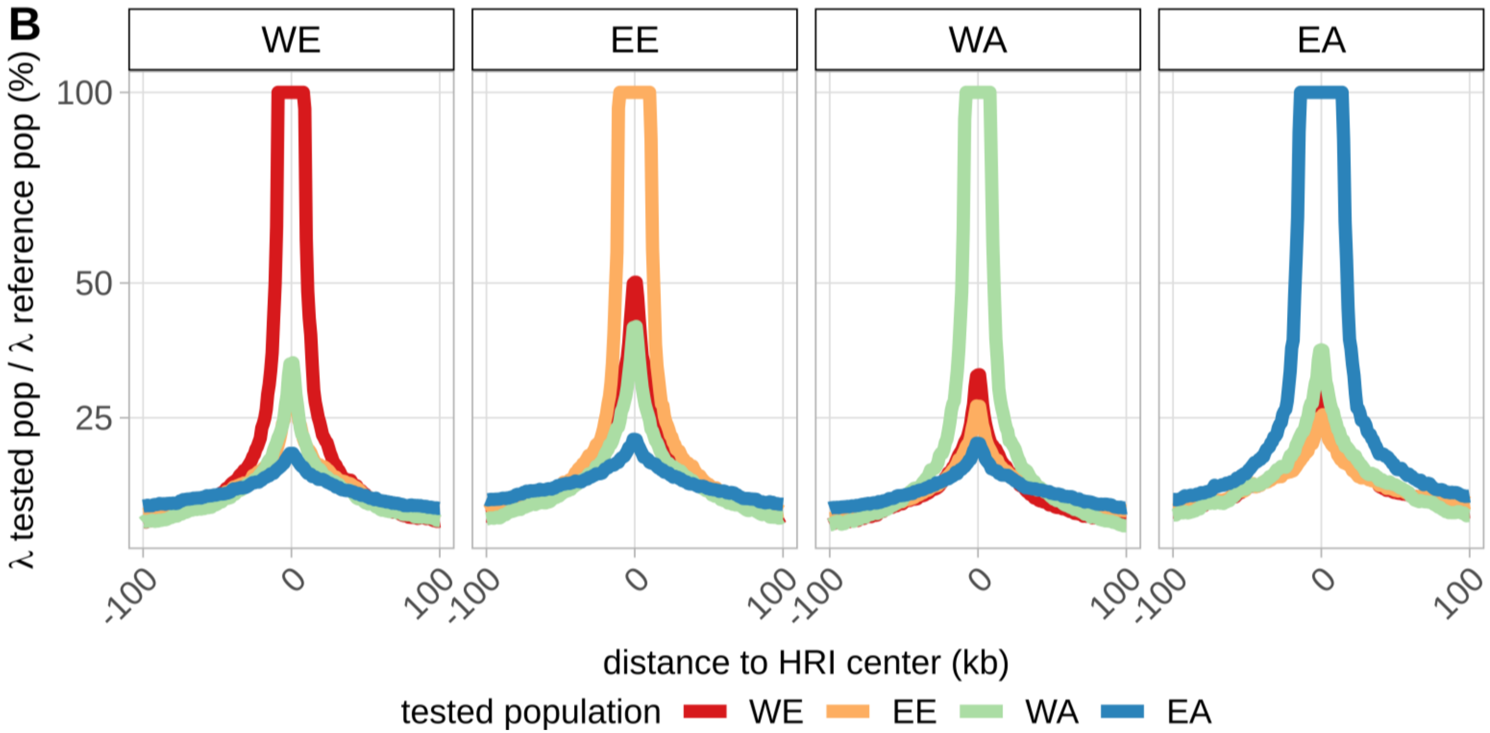
Conservation of HRIs across the 4 populations



- Less population specific intervals than expected
- 34% of HRIs are shared > than expected by chance

Common mechanism driving HRIs positions
but differenciation

Conservation of HRIs across the 4 populations

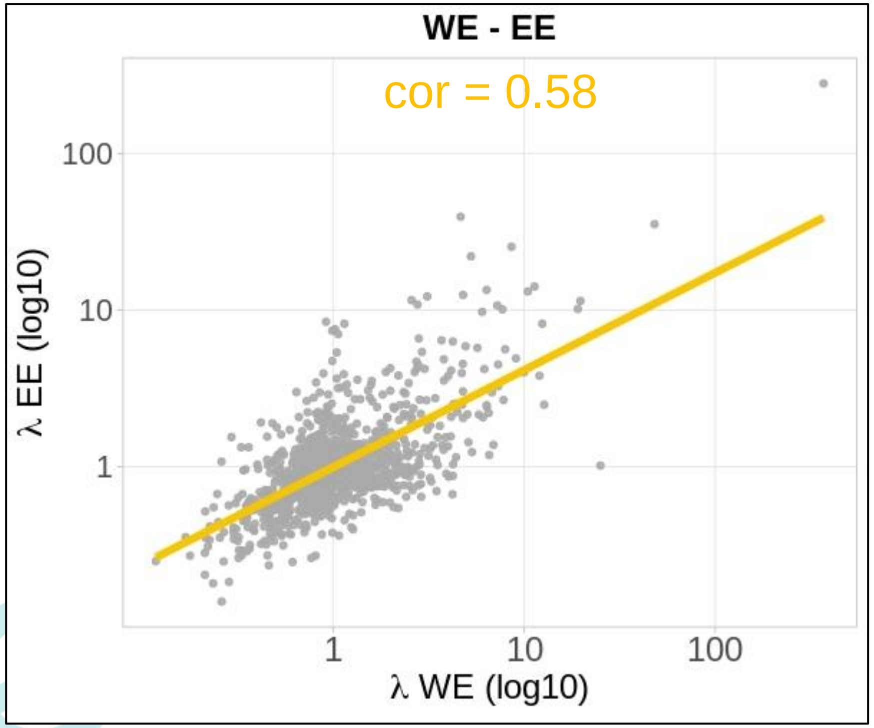


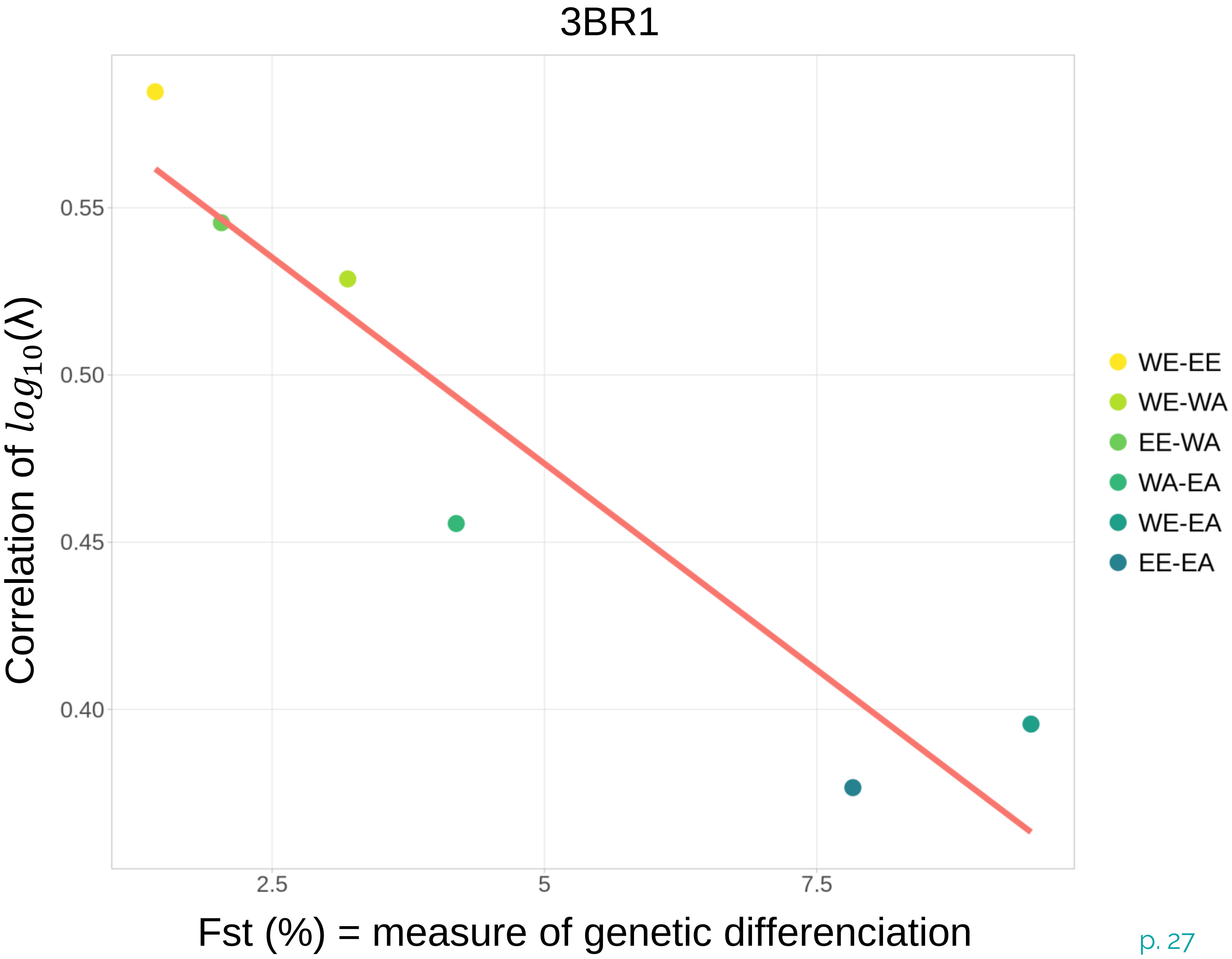
Recombination at HRIs increased in closer populations

Variability of global recombination landscape

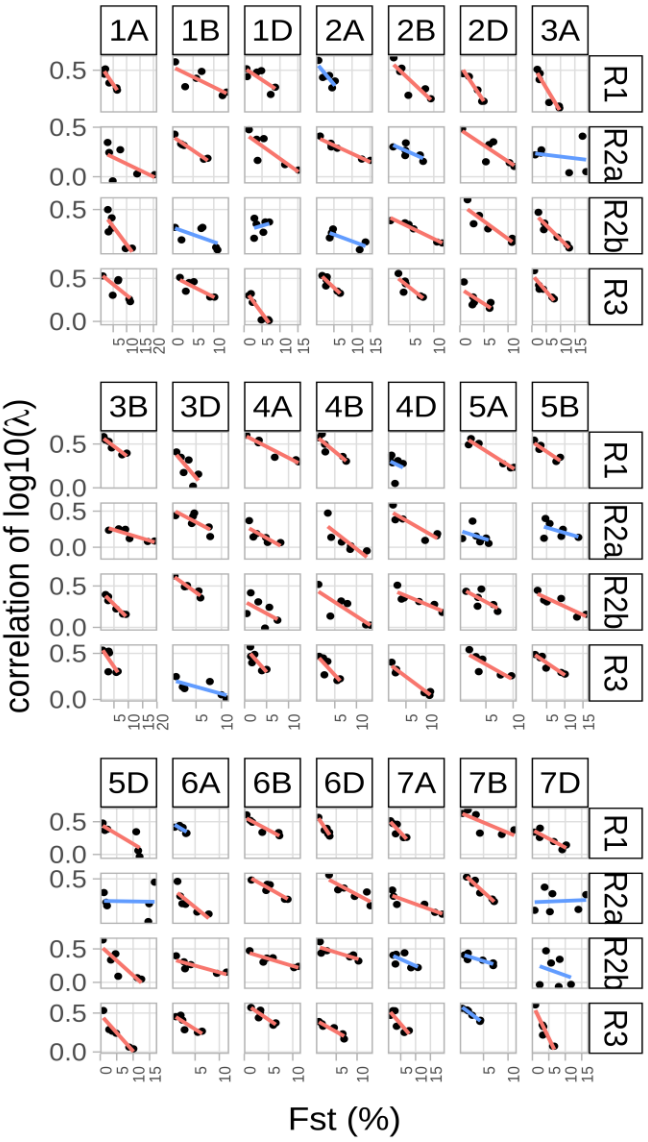
For each genomic region 1AR1 ... 7DR3 and each pair of populations:

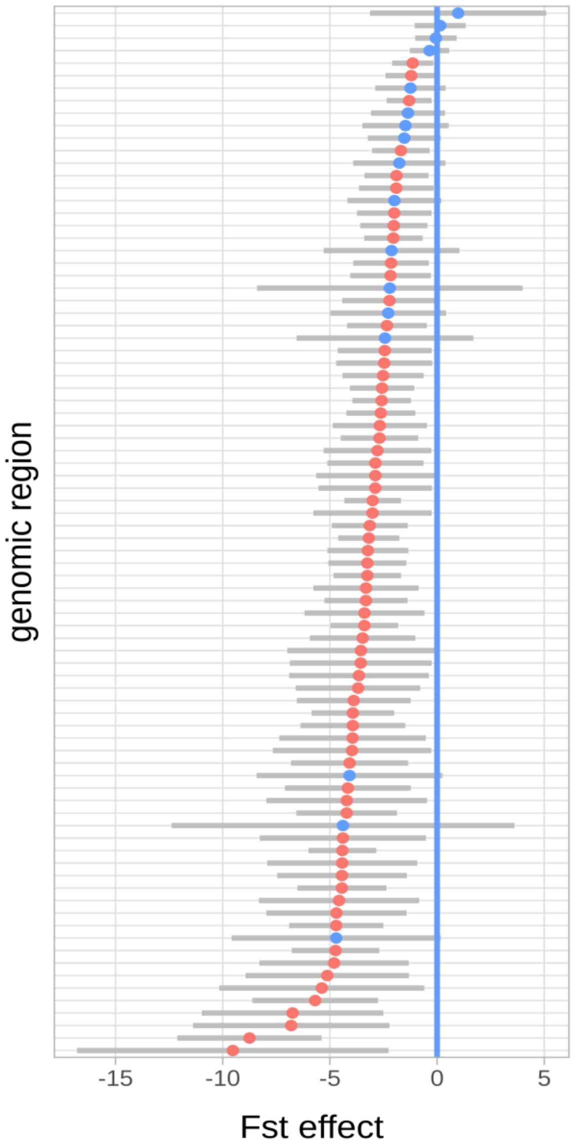
- Compute Fst
- Compute correlation of rec. inflations (λ)





Variability of global recombination landscape





95% confidence interval of slope estimate: 🟥 <0 🟦 include 0

- Negative slopes for most regions
- Most slopes significantly < 0

→ Similary of recombination landscapes decreases with genetic differenciation

Conclusion

Recombination landscape evolved quickly in bread wheat

Genetic divergence associated with increased differenciation of LD patterns

- Among the highest resolutive recombination landscapes of bread wheat
- Very clear signal
- Results likely not biased by evolutionnary forces
- Also observed in other studies on plants:
  - Bread wheat (Gardiner, 2019, Genome Biology)
  - Rice (Marand et al. 2019, The Plant Cell)
  - Cacao (Schwarskops et al. 2020, BMC genomics)

+ Maize (Rodgers-Melnick et al. 2015, PNAS), Poplar (Wang et al. 2016, Genetics), Cotton (Shen et al. 2019, The Plant journal)

→ Evolution of CO repartition along the genome in plants

Discussion: Drivers of the evolution of CO repartition along the genome

→ Might provide some insights about CO position determinism

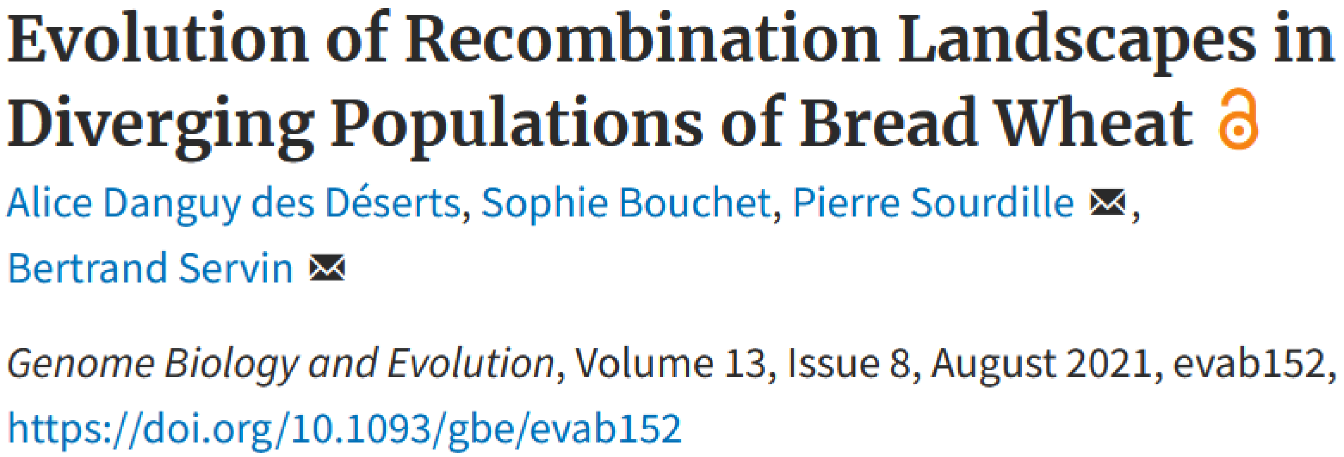
How to explain such variability ?

- Environmental effects: T°C
- (Epi-)Genomic variability: DNA sequence or chromatine landscape
- Genes driving CO position (Ex: PRDM9 in few mammals)

Perspectives:

- Detect recombination hotspots
- Functionnal annotation + chromatine marks at hotspots
- GWAS in segregating families

Thank you for your attention



Evolution of Recombination Landscapes in Diverging Populations of Bread Wheat

Alice Danguy des Déserts, Sophie Bouchet, Pierre Sourdille ✉,
Bertrand Servin ✉

Genome Biology and Evolution, Volume 13, Issue 8, August 2021, evab152,
https://doi.org/10.1093/gbe/evab152

Corresponding authors ✉
Pierre Sourdille: pierre.sourdille@inrae.fr
Bertrand Servin: bertrand.servin@inrae.fr