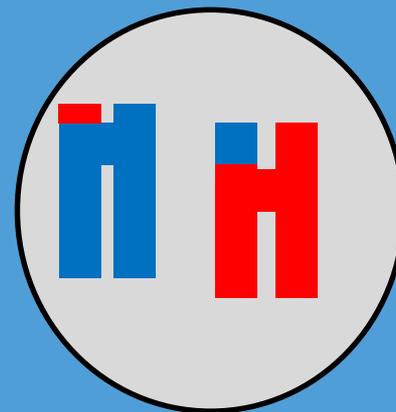
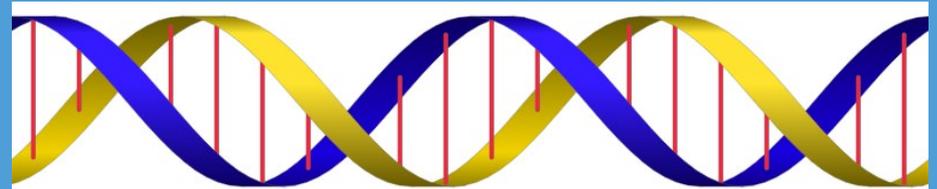
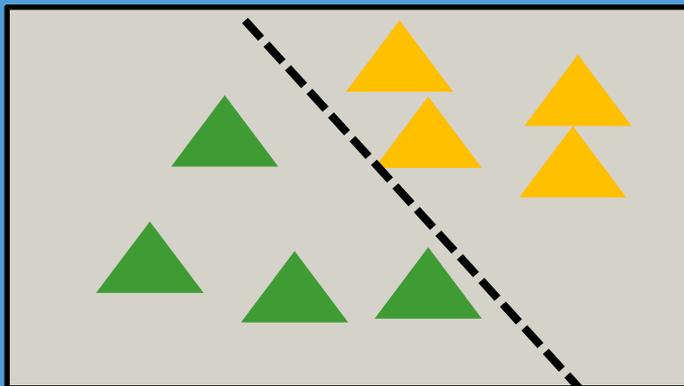


Machine learning based prediction and analysis of meiotic crossovers

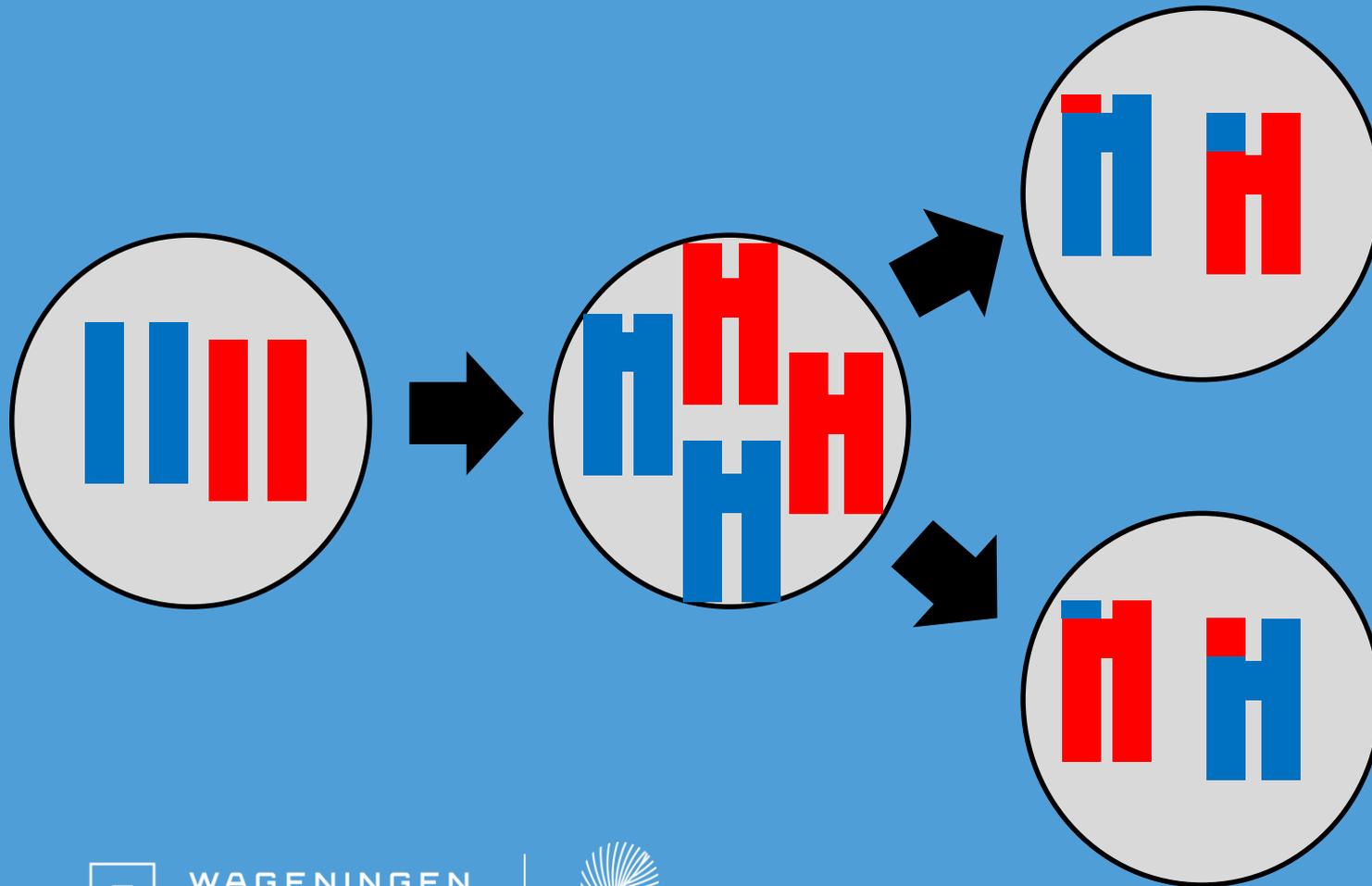
Aalt-Jan van Dijk

Bioinformatics & Biometris

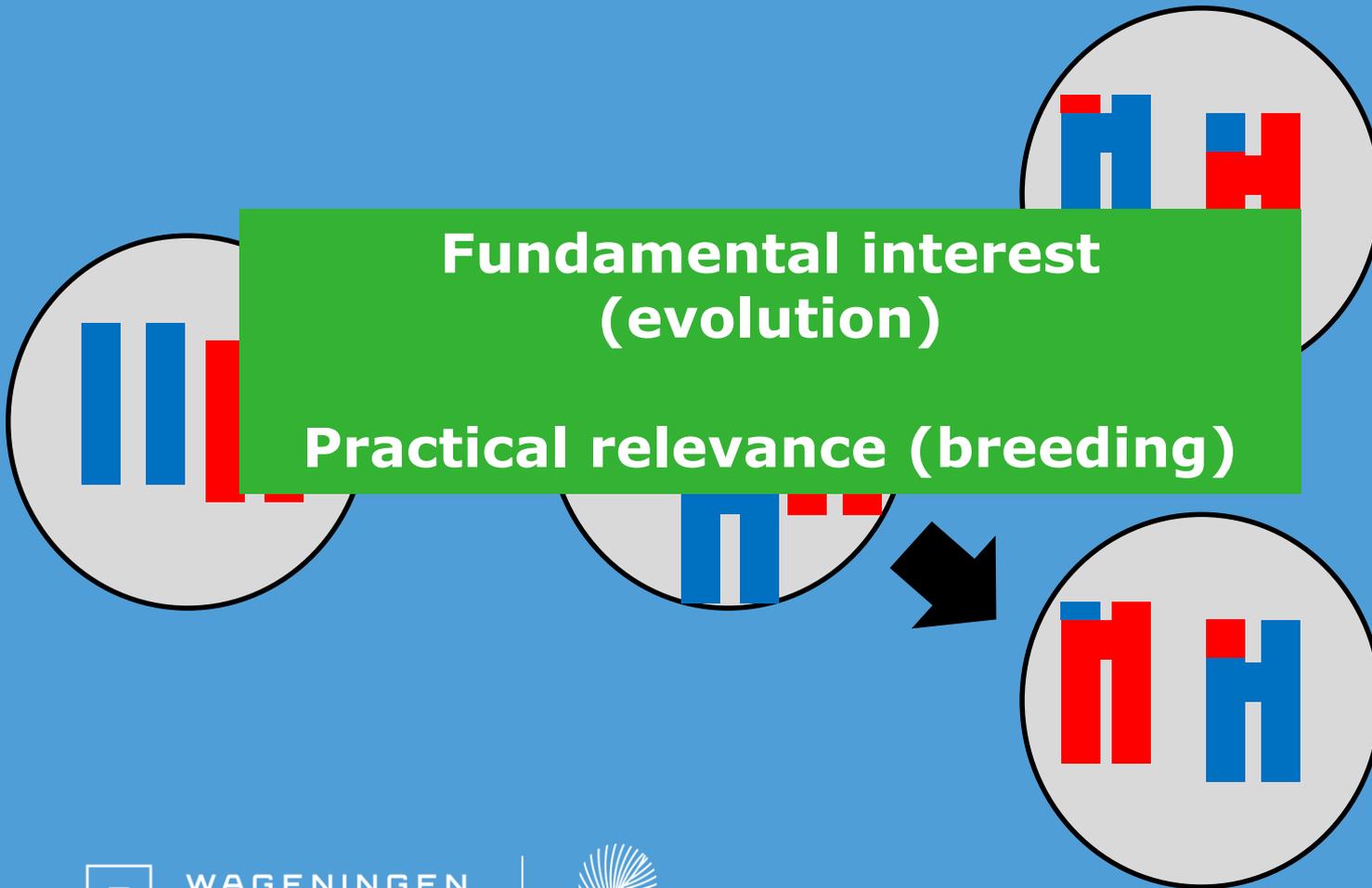
Wageningen University



Meiotic crossovers



Meiotic crossovers



More and more data on crossovers in plants

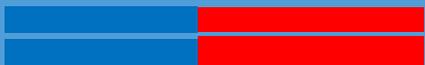
Double Strand Break



Single end invasion



....



RESEARCH

Open Access

Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency



Single gametophyte sequencing reveals that crossover events differ between sexes in maize

Cheng Luo¹, Xiang Li¹, Qinghua Zhang¹ & Jianbing Yan¹

Meiotic crossovers are associated with open chromatin and enriched with *Stowaway* transposons in potato

Alexandre P. Marand, Shelley H. Jansky, Hainan Zhao, Courtney P. Leisner, Xiaobiao Zhu, Zixian Zeng, Emily Crisovan, Linsey Newton, Andy J. Hamernik, Richard E. Veilleux, C. Robin Buell and Jiming Jiang

Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*

Sevgin Demirci, Aalt D.J. van Dijk, Gabino Sanchez Perez, Saulo A. Aflitos, Dick de Ridder, Sander A. Peters



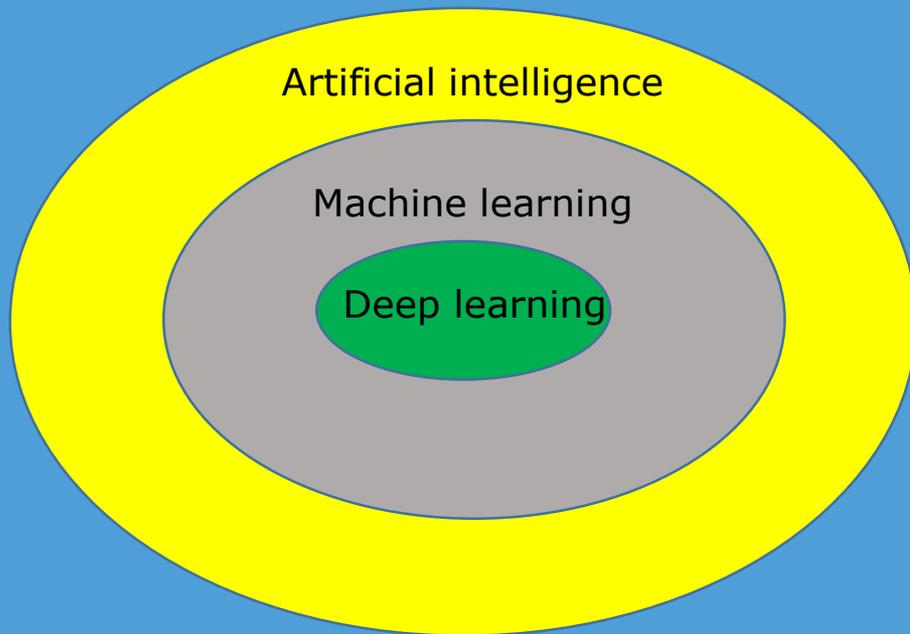
WAGENINGEN
UNIVERSITY & RESEARCH



100years
1918 — 2018

Demirci Plant Journal 2017

Goal: get insight using machine learning

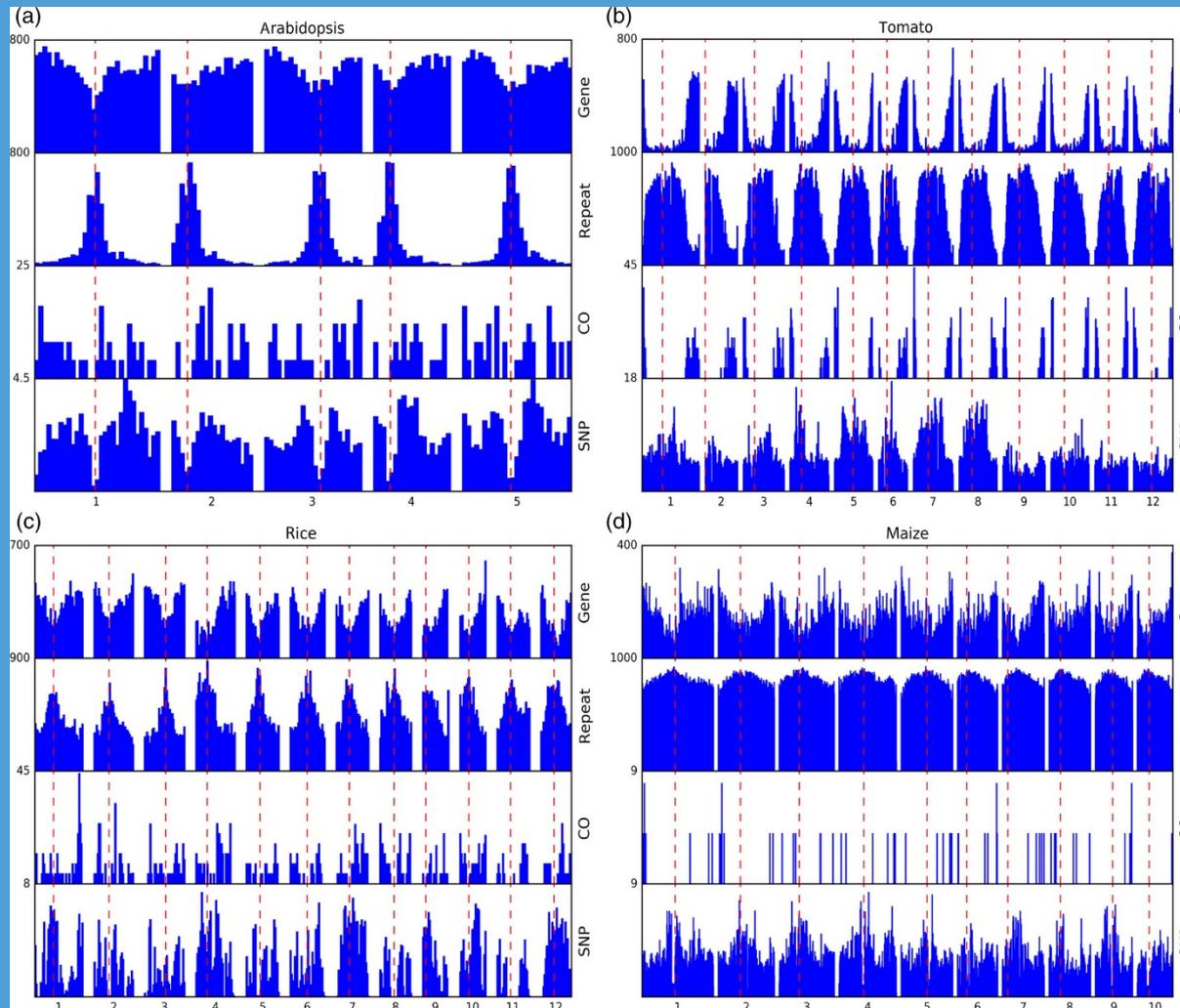


Machine learning:

- use **training set**
- find **patterns** in data
- predict on **test set**
- **interpret** the model



Datasets in four plant species

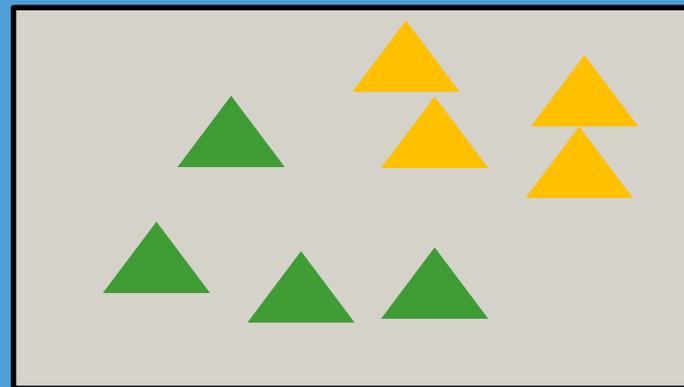


Crossover datasets

Training set: 'positive cases' and 'negative cases'

	Tomato	Arabidopsis	Rice	Maize
Number of 4kb regions containing a crossovers	664	159	468	63

ML then learns to discriminate positives and negatives

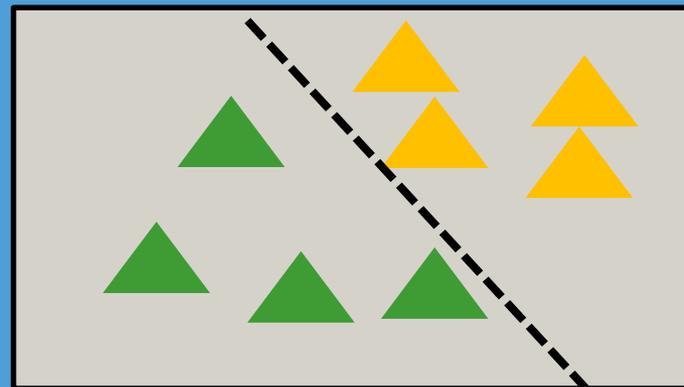


Crossover datasets

Training set: 'positive cases' and 'negative cases'

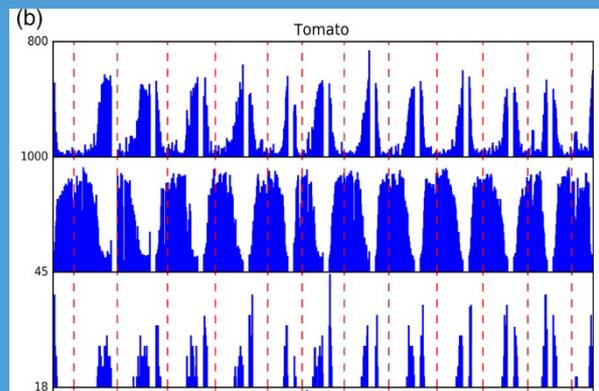
	Tomato	Arabidopsis	Rice	Maize
Number of 4kb regions containing a crossovers	664	159	468	63

ML then learns to discriminate positives and negatives



Sampling negative cases from....

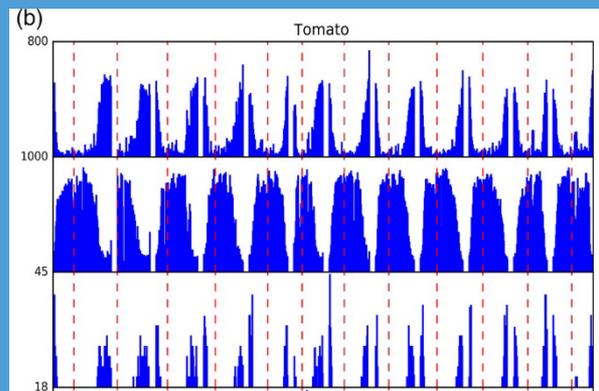
Whole genome → model learns mainly to discriminate euchromatin from heterochromatin



Sampling negative cases from....

~~Whole genome~~ → ~~model learns mainly to discriminate~~
~~euchromatin from heterochromatin~~

Euchromatin only → ignores crossovers in heterochromatin

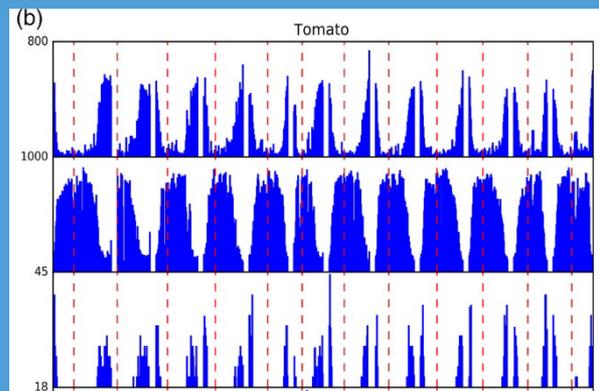


Sampling negative cases from....

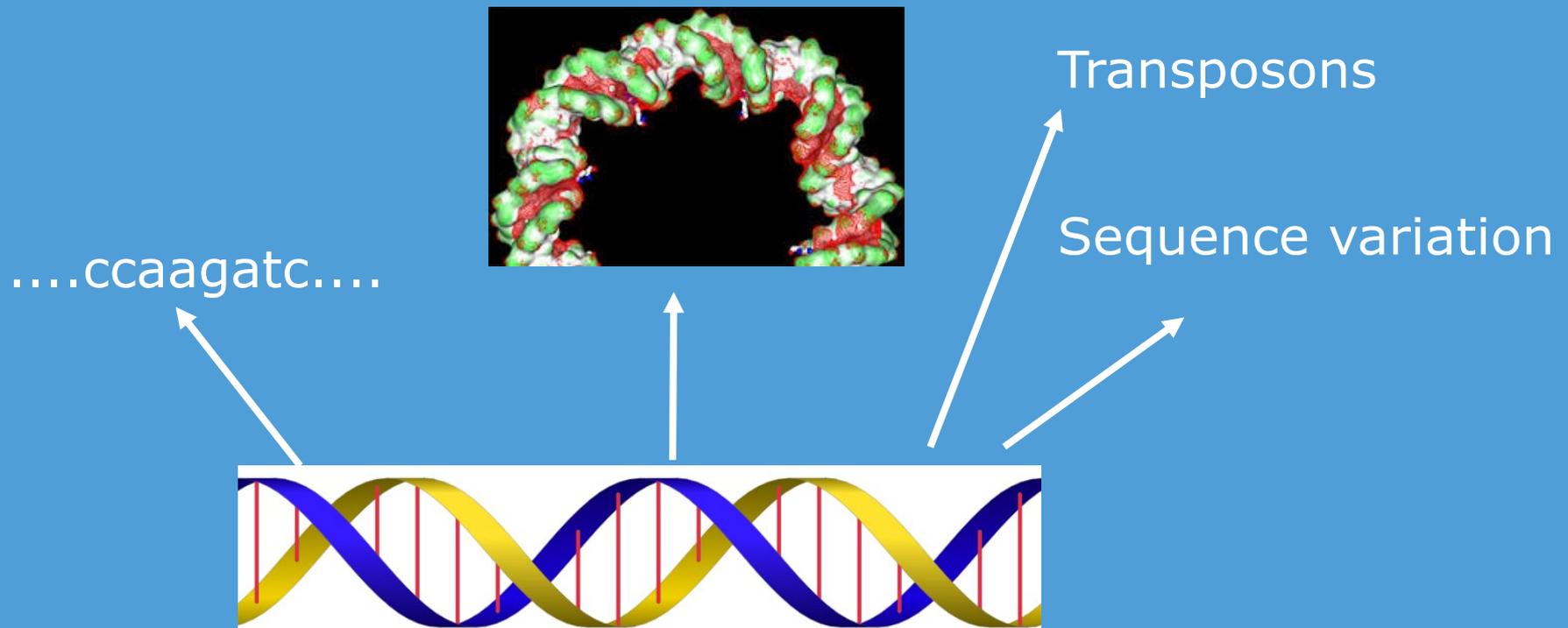
~~Whole genome~~ → ~~model learns mainly to discriminate~~
~~euchromatin from heterochromatin~~

~~Euchromatin only~~ → ~~ignores crossovers in heterochromatin~~

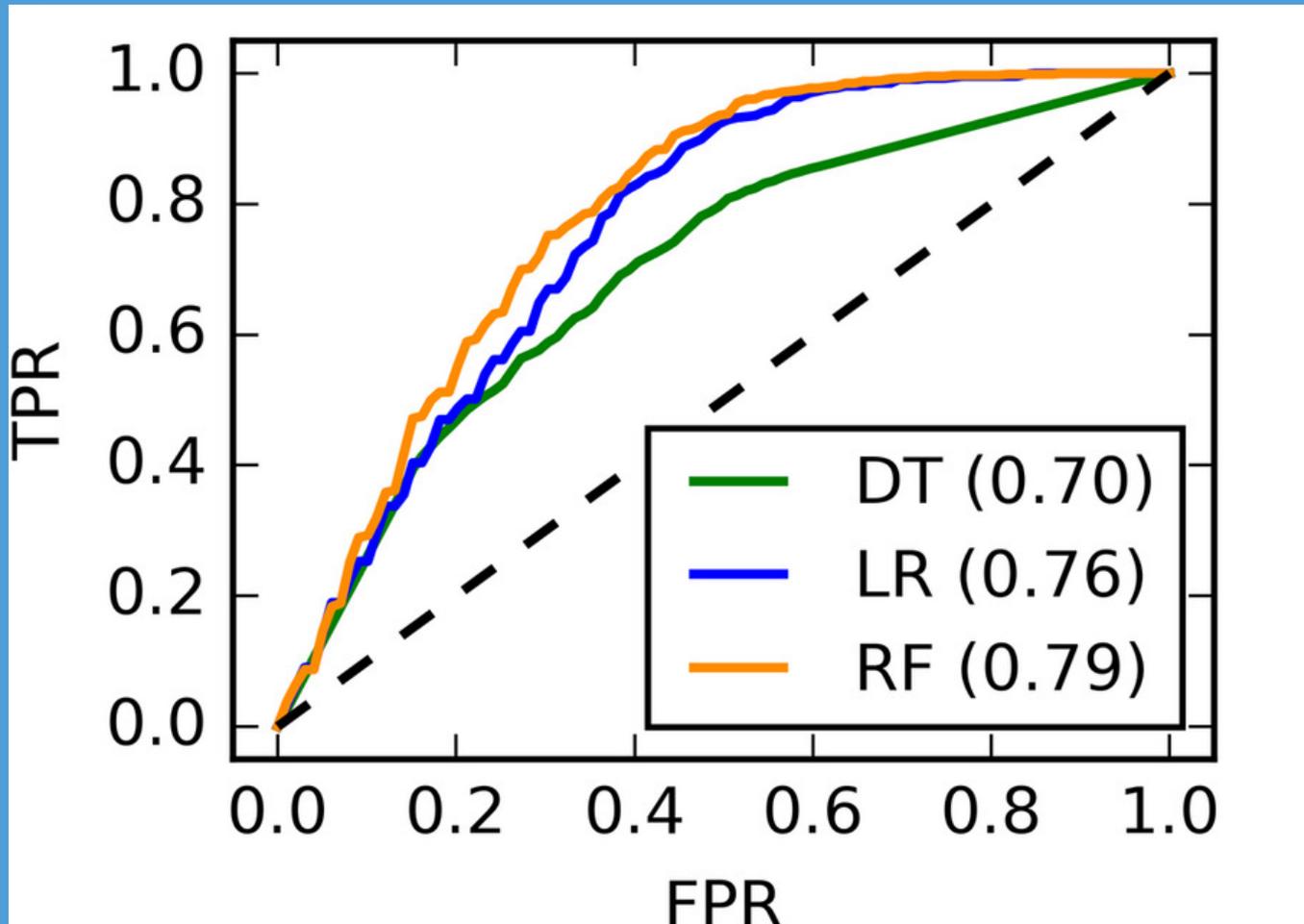
Based on gene density → model learns to predict for
a given gene-rich region whether crossovers are likely or not



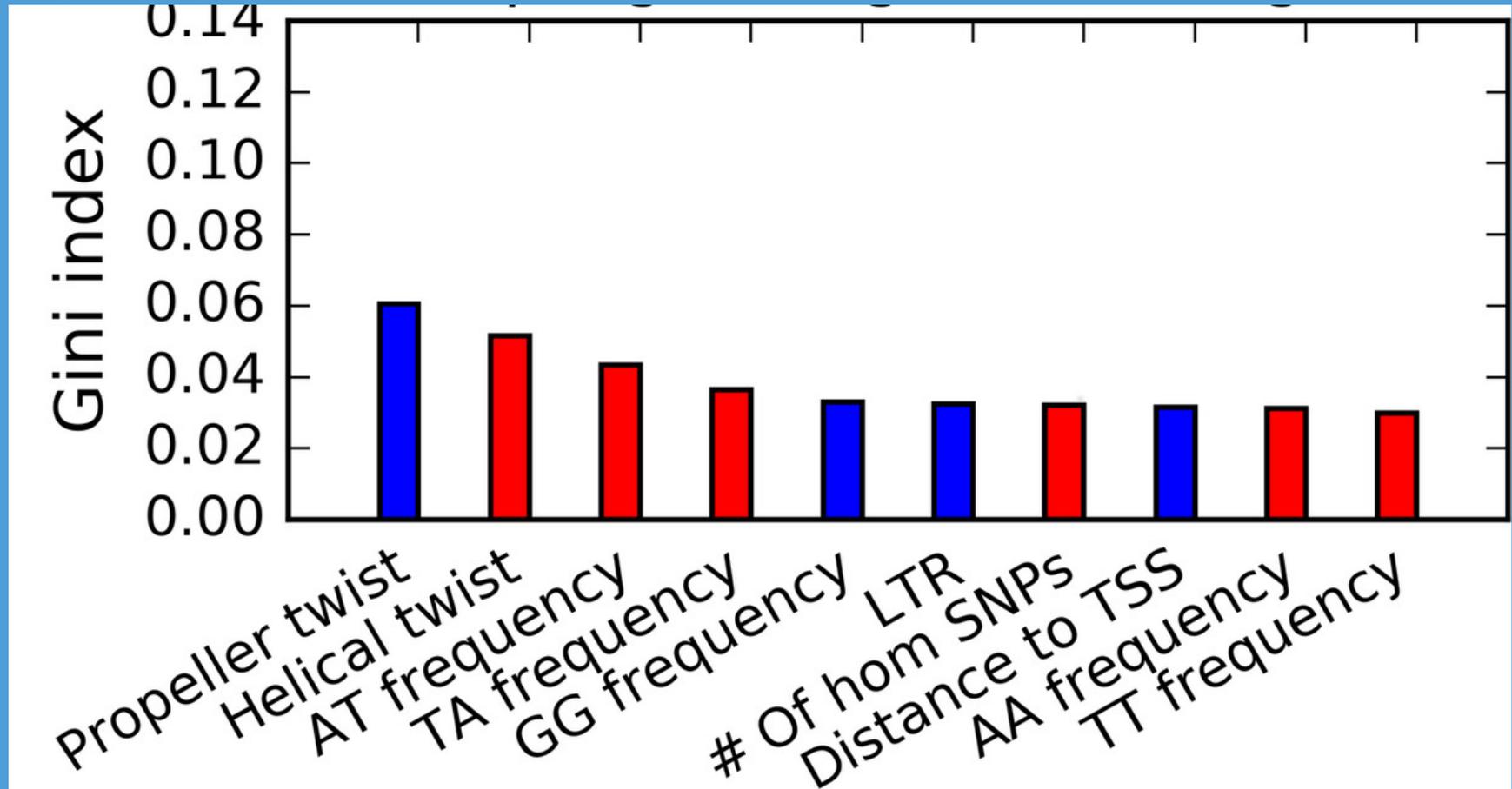
Features



Results for tomato: prediction performance



Results for tomato: important features

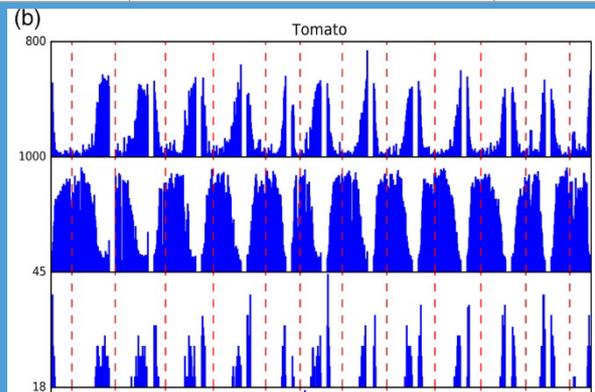


Prediction performance four species

	Tomato	Arabidopsis	Rice	Maize
AUROC	0.79 (0.04)	0.63 (0.08)	0.67 (0.05)	0.72 (0.14)

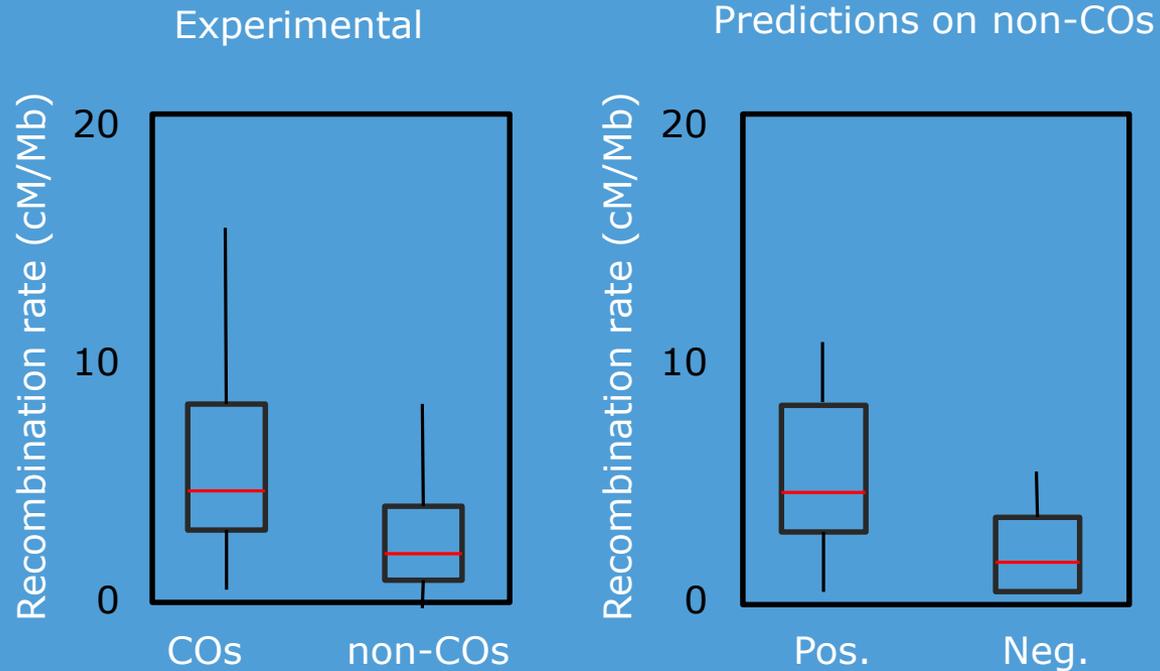
Additional validation (I)

	Tomato	Arabidopsis	Rice	Maize
AUROC	0.79 (0.04)	0.63 (0.08)	0.67 (0.05)	0.72 (0.14)
Pericentromeric regions	0.95	0.66	0.76	0.92

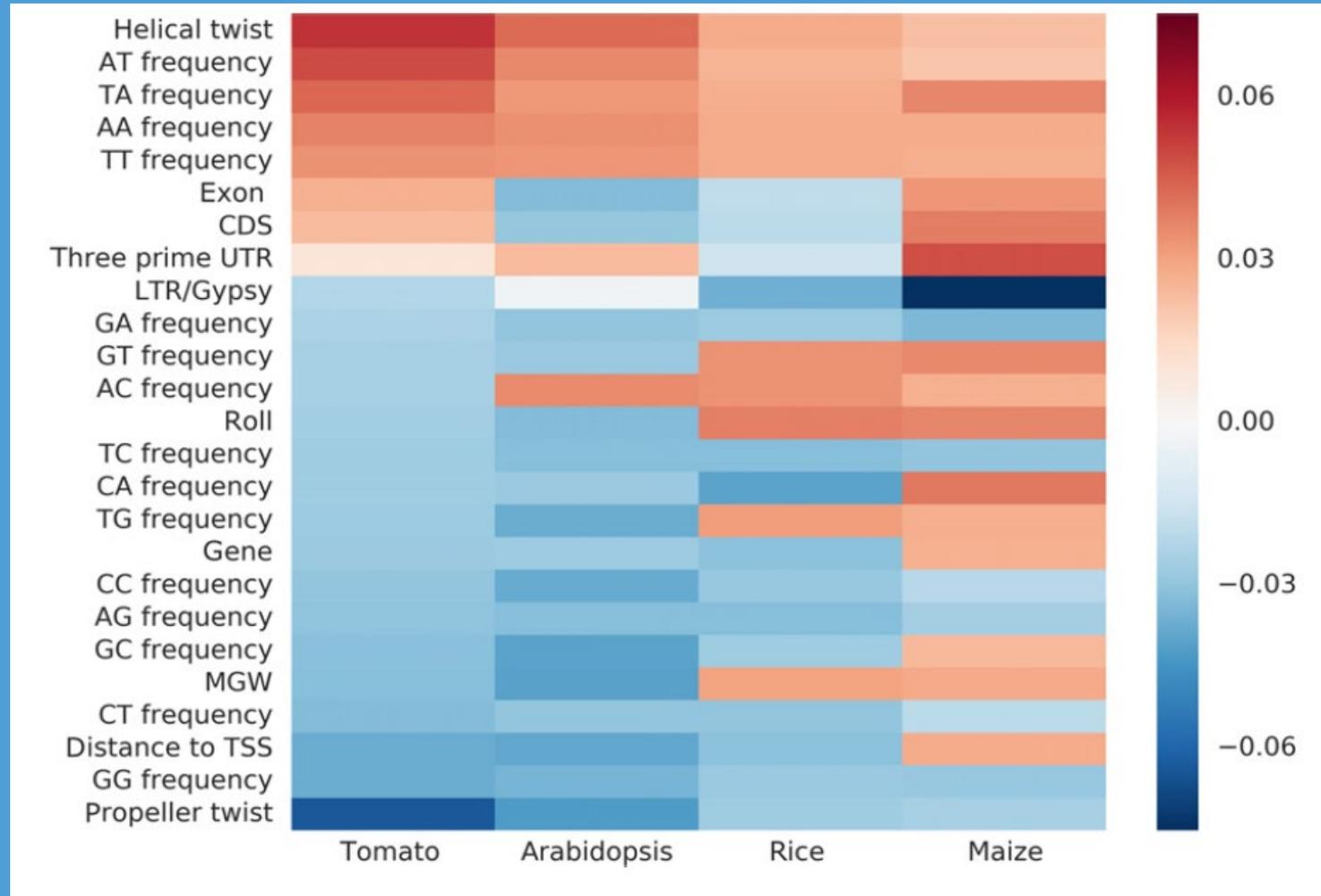


Additional validation (II)

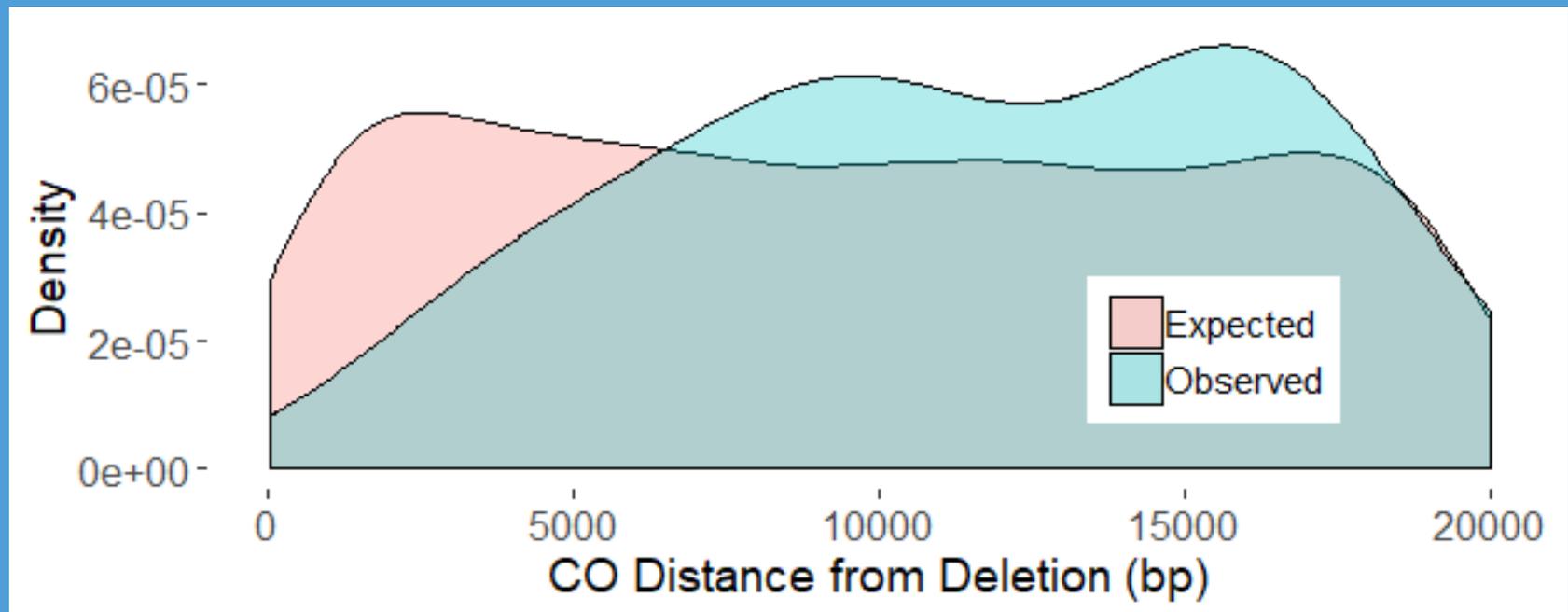
Historical recombination rate (Arabidopsis)



Four species: similarities and differences



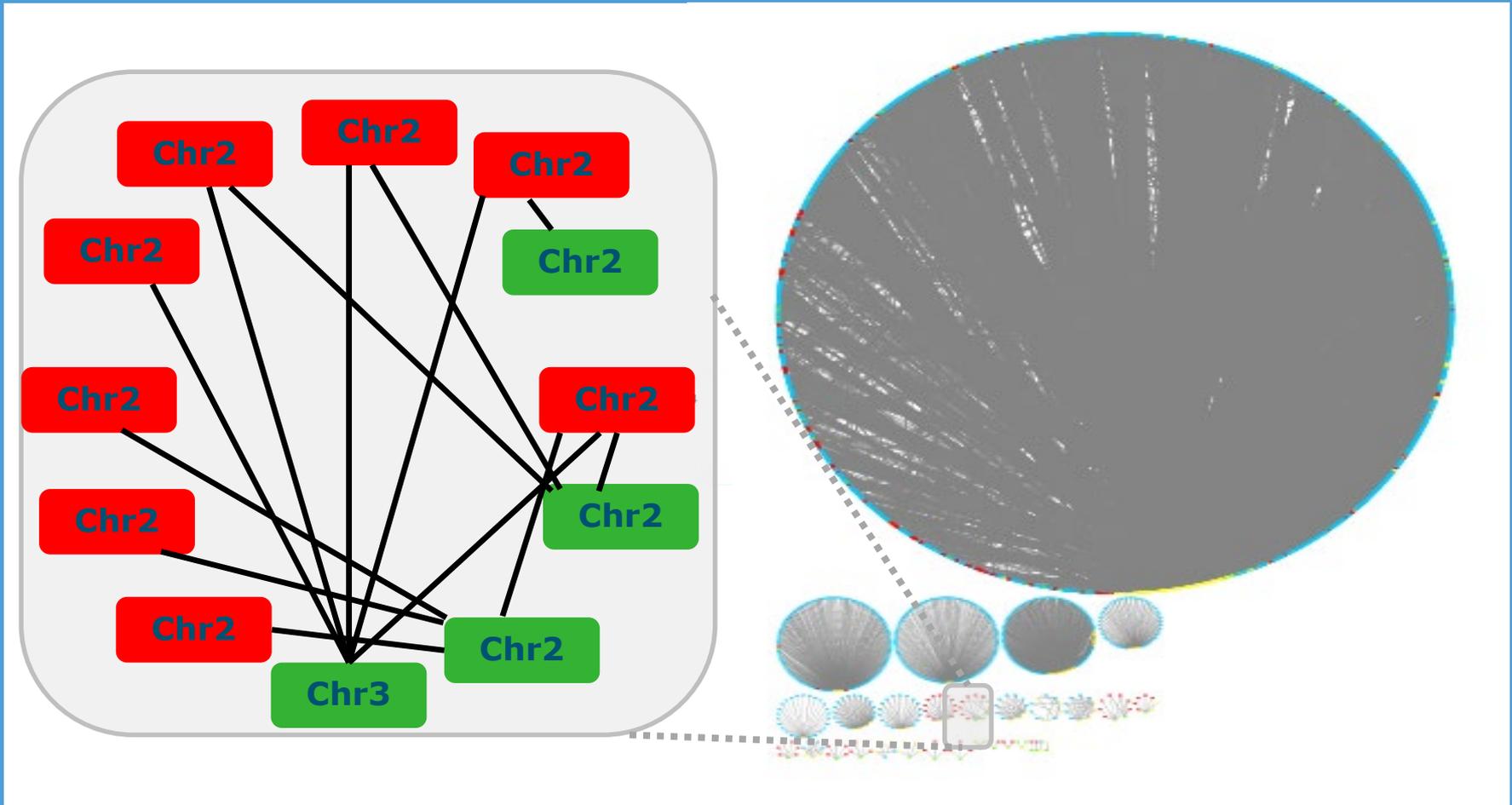
Next (1): Structural variants



rice

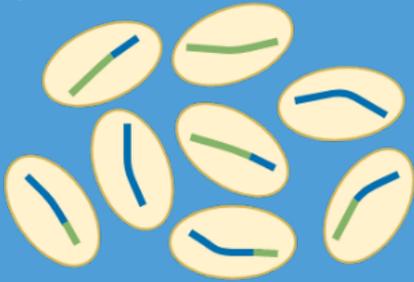


Next (2): crossovers vs synteny blocks



Next (3): More data

F1 pollen

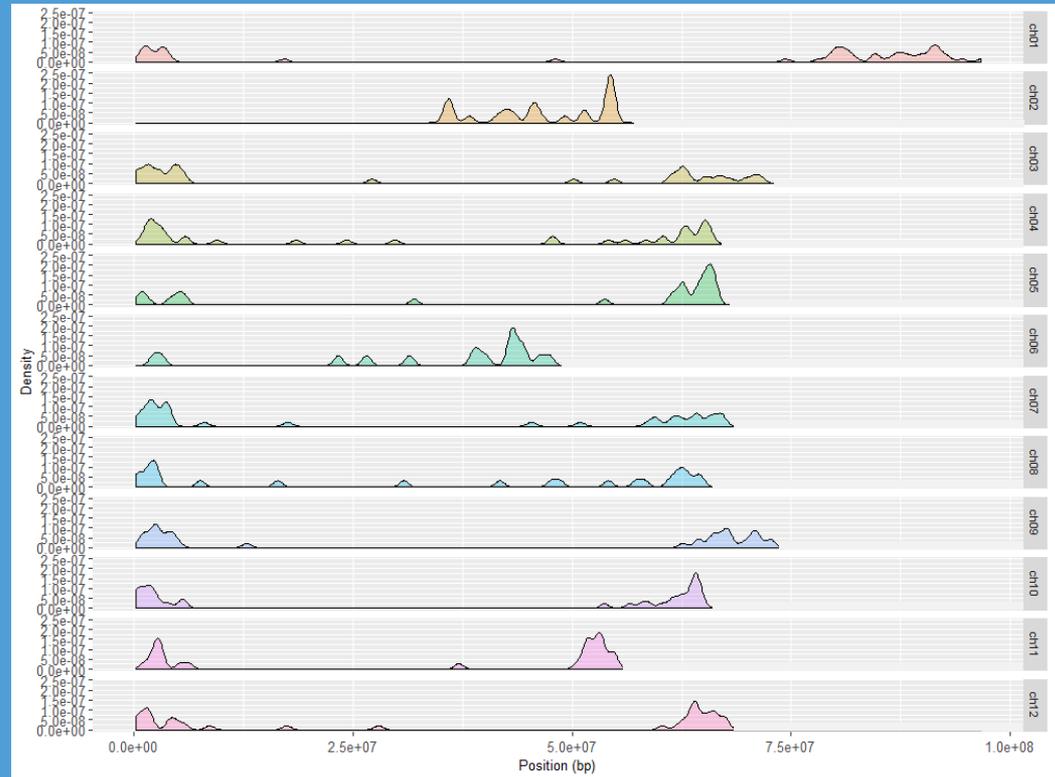
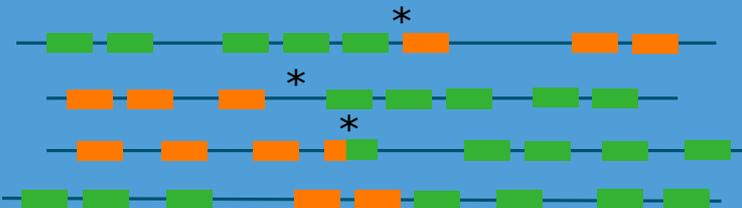


10x Genomics



HiSeq2500

Crossover detection



Conclusions

Prediction of crossovers

Insight into underlying determinants

Soon: more data and more models



Acknowledgements

Wageningen University and Research:

Sander Peters

Dick de Ridder

Sevgin Demirci

Roven Rommel Fuentes

Funding:



WAGENINGEN
UNIVERSITY & RESEARCH



100years
1918 — 2018