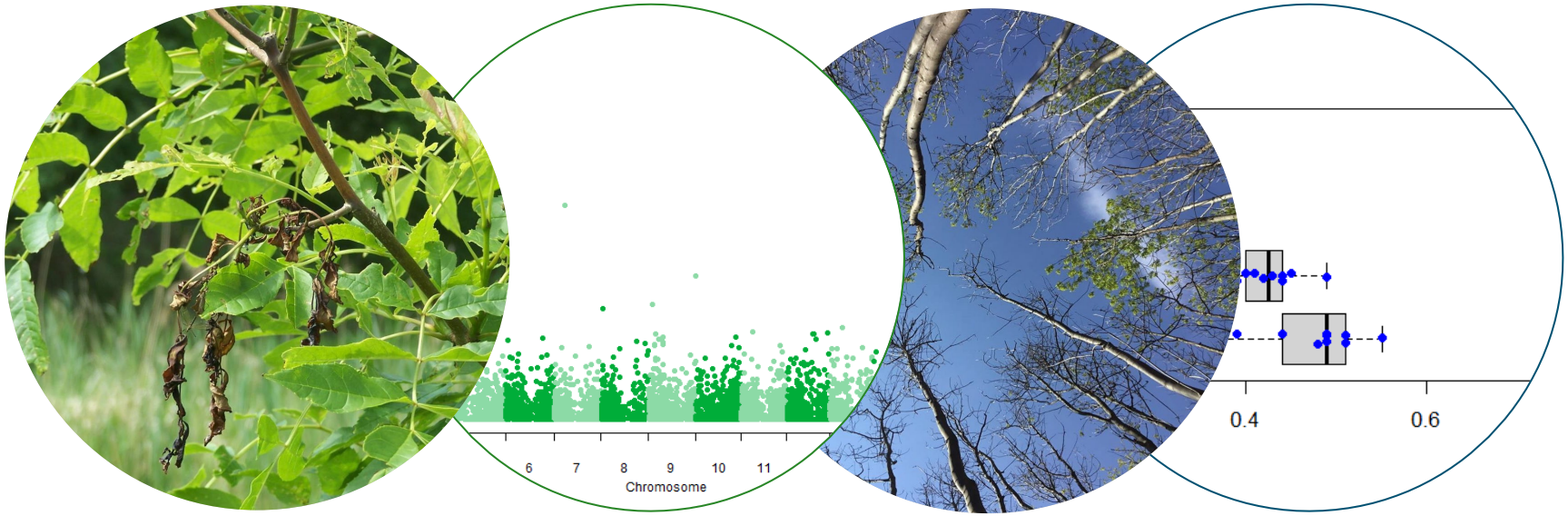
In search for resistance to ash dieback disease in *F. excelsior* using GWAS and genomic prediction analyses



WAGENINGEN
UNIVERSITY & RESEARCH

Giorgio Tumino, 20-06-2022

Content

Genotypes
- 2000 DNA samples (WUR + INRAE)
- SNP arrays (pre-screening and 4TREE arrays)

Population structure and relatedness

Linkage Disequilibrium (LD) decay

Ash dieback phenotypes

GWAS and genomic prediction

Conclusions

Genotype datasets

4TREE SNP array

1. **1704 SNPs** associated with traits of interest (1137 dieback, incompatibility and ash borer)
2. **197 SNPs** selected to discriminate between F. angustifolia and F. excelsior
3. **9698 SNPs** chosen to cover the F. excelsior genome
4. **1808** additional high-quality SNPs

13407 SNPs

SNP positions based on a recent assembly by a Poland group (Kazimierz Wielki University).

91% of 4TREE SNPs were uniquely located on one of the 23 chromosomes.

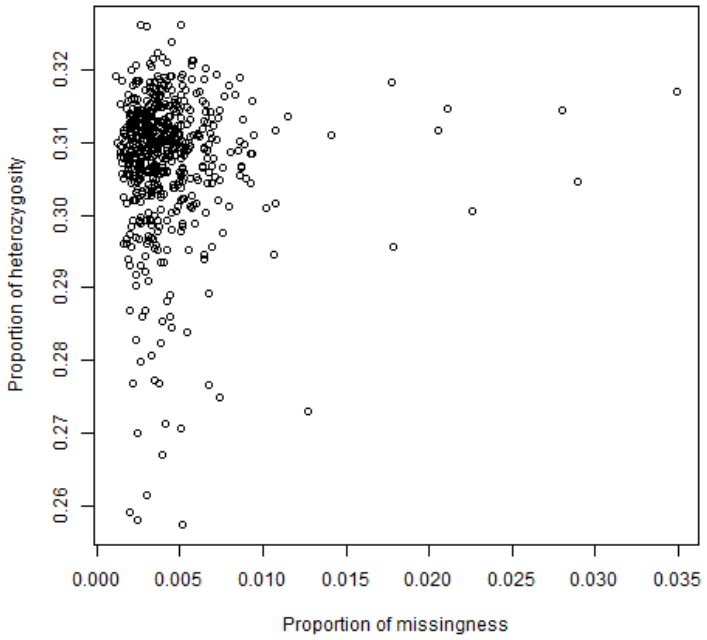
Genotype datasets

Number of samples (after quality tests)

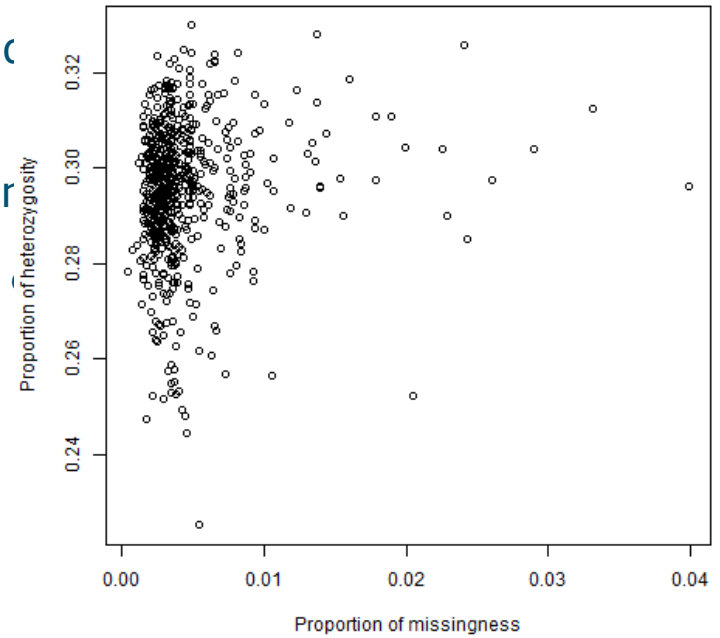
	Pre-screening 138 K SNPs	4TREE 13 K ash SNPs
WUR	70	899
INRAE	57	976

2002 samples

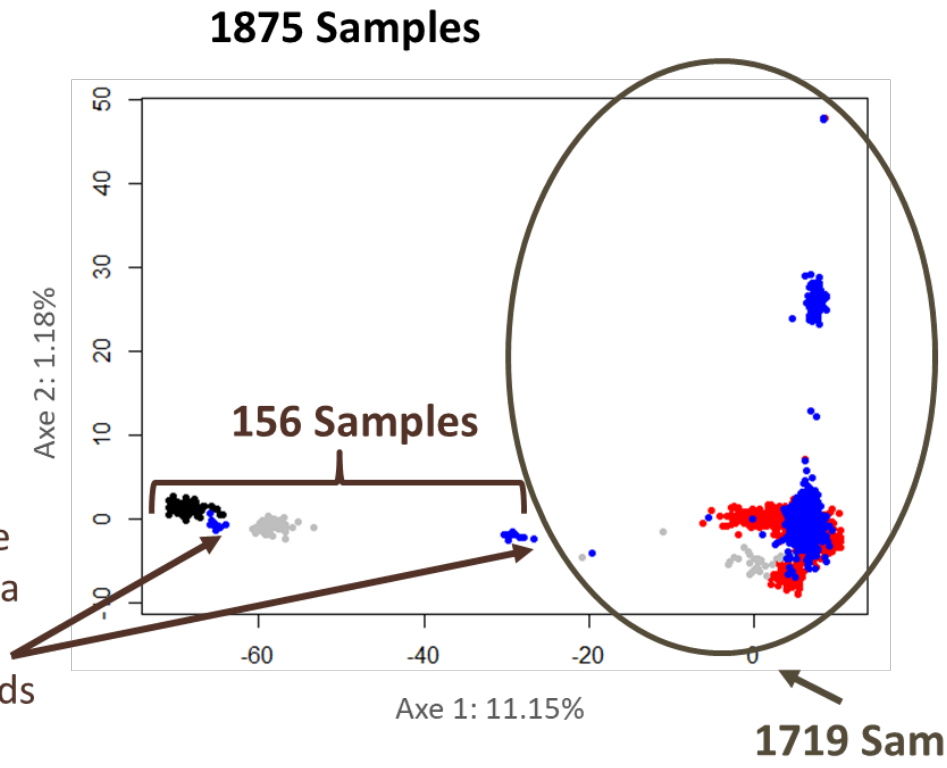
WUR Windesheim trial (600 samples)



INRAE Moulin-sur-Yèvre trial (600 samples)



Population structure



- 18 samples with the same origin code, from Romania
- 1 sample from Netherlands

F. angustifolia (INRAE)

Plantation with a large phenotypic diversity (INRAE)

F. excelsior (INRAE)

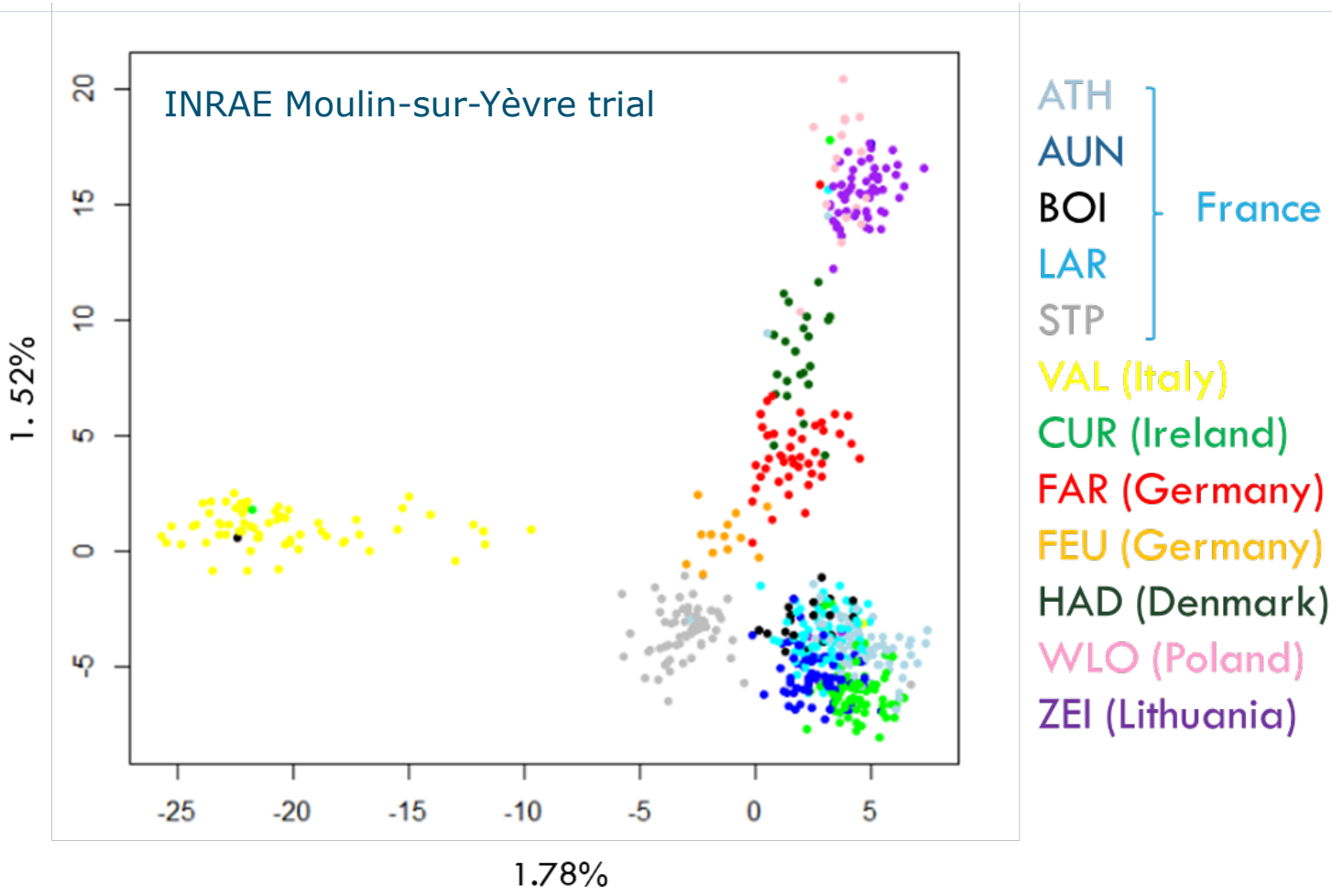
F. excelsior (WUR)

by Sarah Ben-Sadoun

The first coordinate separates *F. angustifolia* and *F. excelsior.*

Potential hybrids have been identified.

Population structure



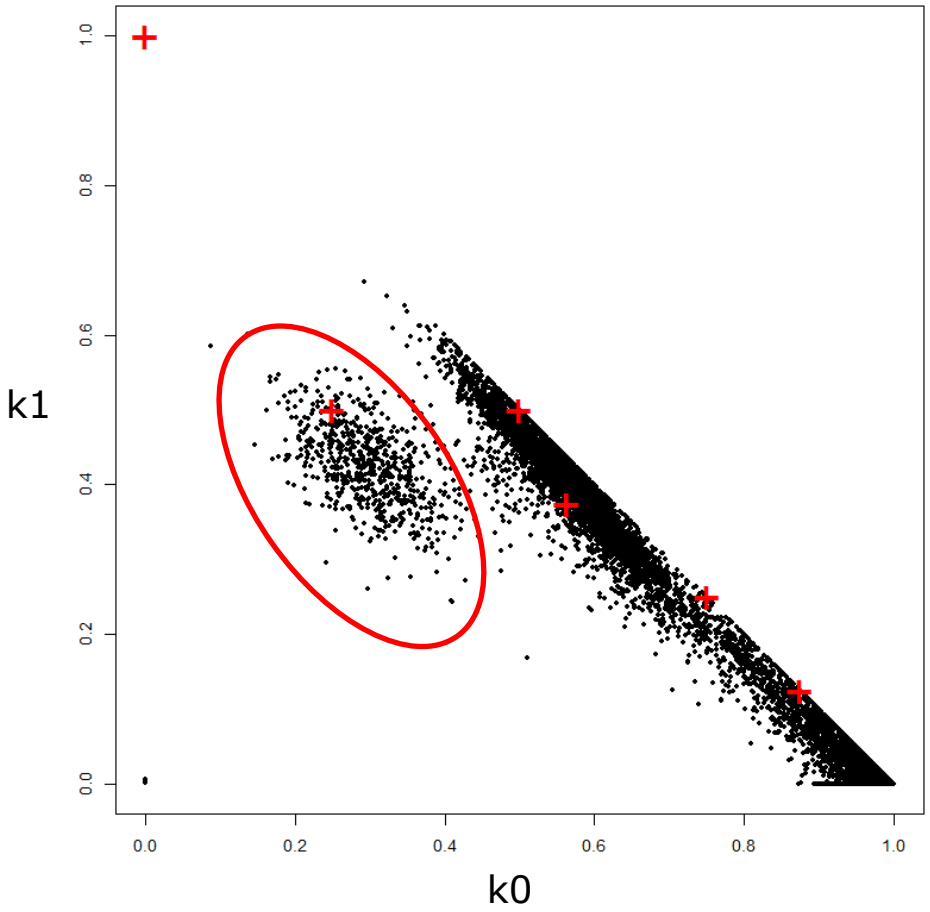
by Sarah Ben-Sadoun

Clustering of samples according to Provenance

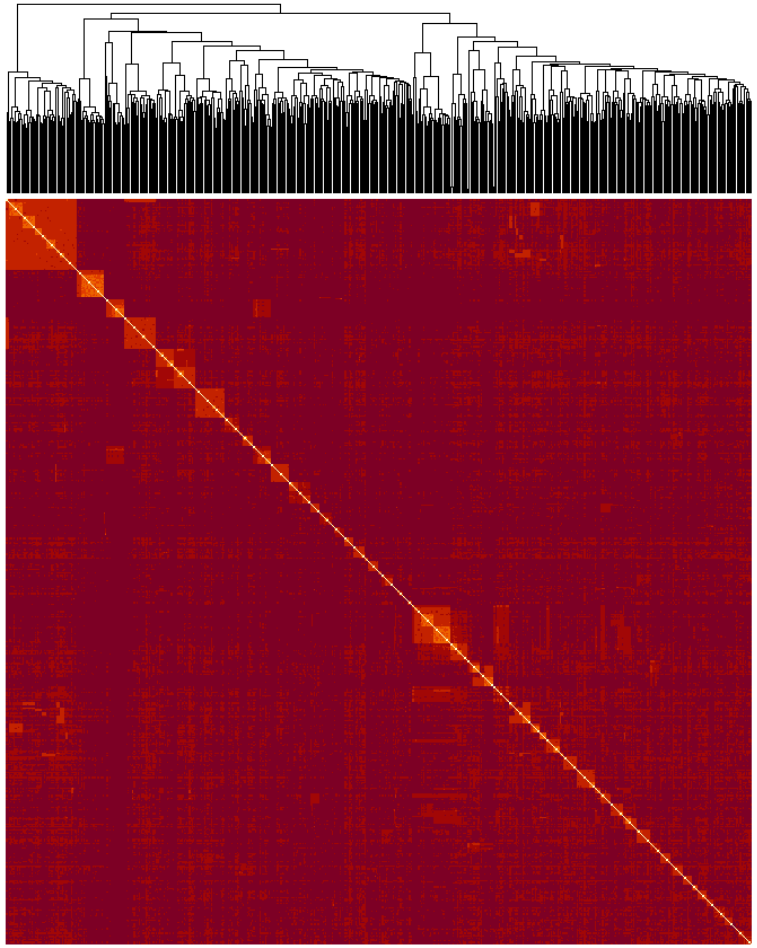
Relatedness

WUR Windesheim trial

IBD coefficients



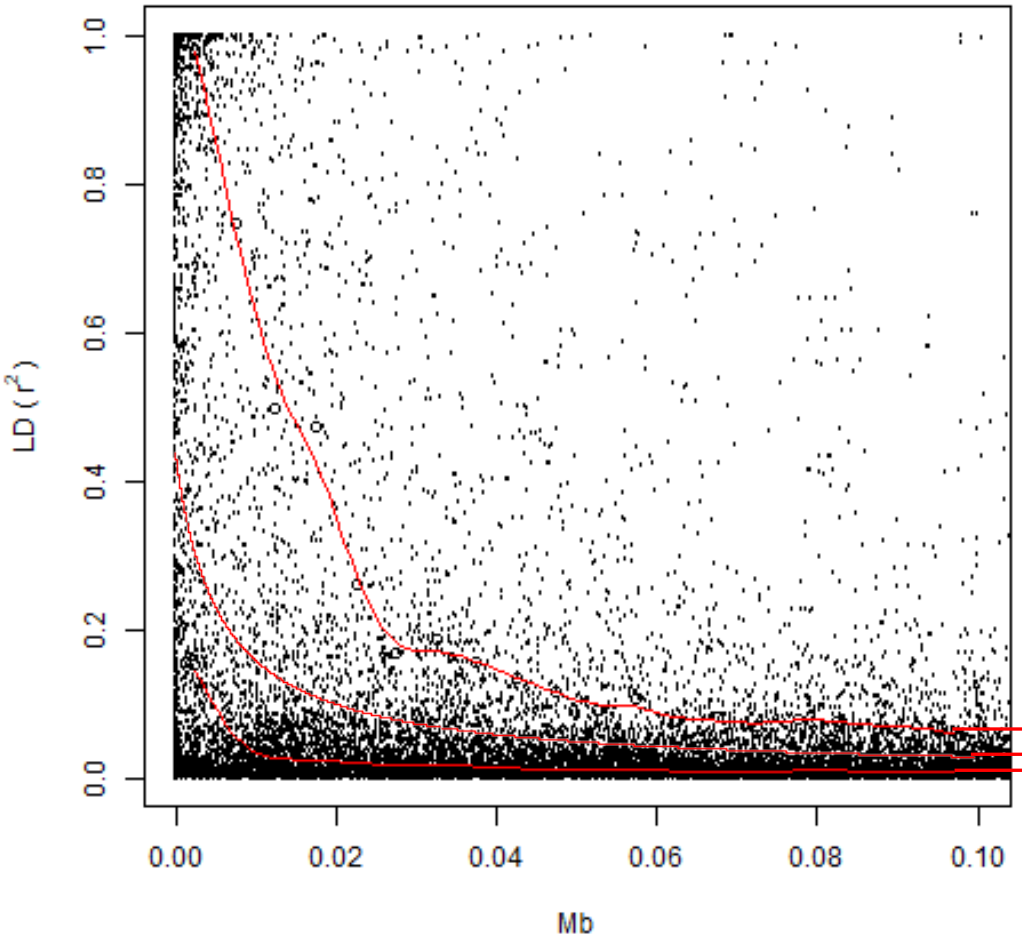
Euclidean distances



Possible to distinguish full-sibs

LD decay

WUR Windesheim trial



LD estimated by r^2

1) Fit percentiles after binning (bin 0.005 Mb)
   - Spline on the 90th percentile
   - Spline on the 50th percentile

2) Fit all points by Hill and Weir curve

90th: $LD_{0.2,90}$ = 26 Kb; $LD_{1/2,90}$ = 15 Kb

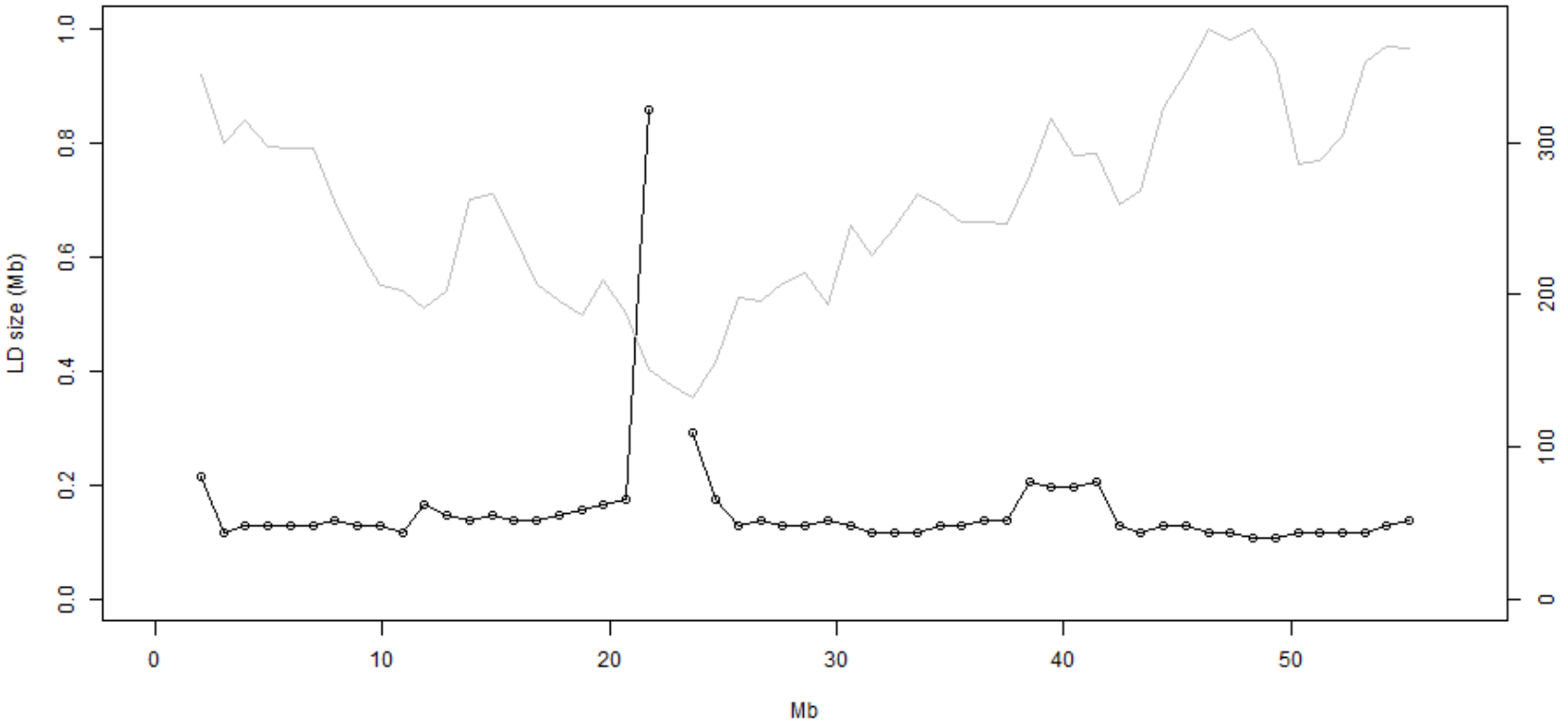
Hill and Weir: $LD_{0.2}$ = 5.5 Kb; $LD_{1/2}$ = 7 Kb

50th: $LD_{1/2,50}$ = 6.4 Kb

LD decay

Pre-screening array (138 K SNPs x 127 samples)

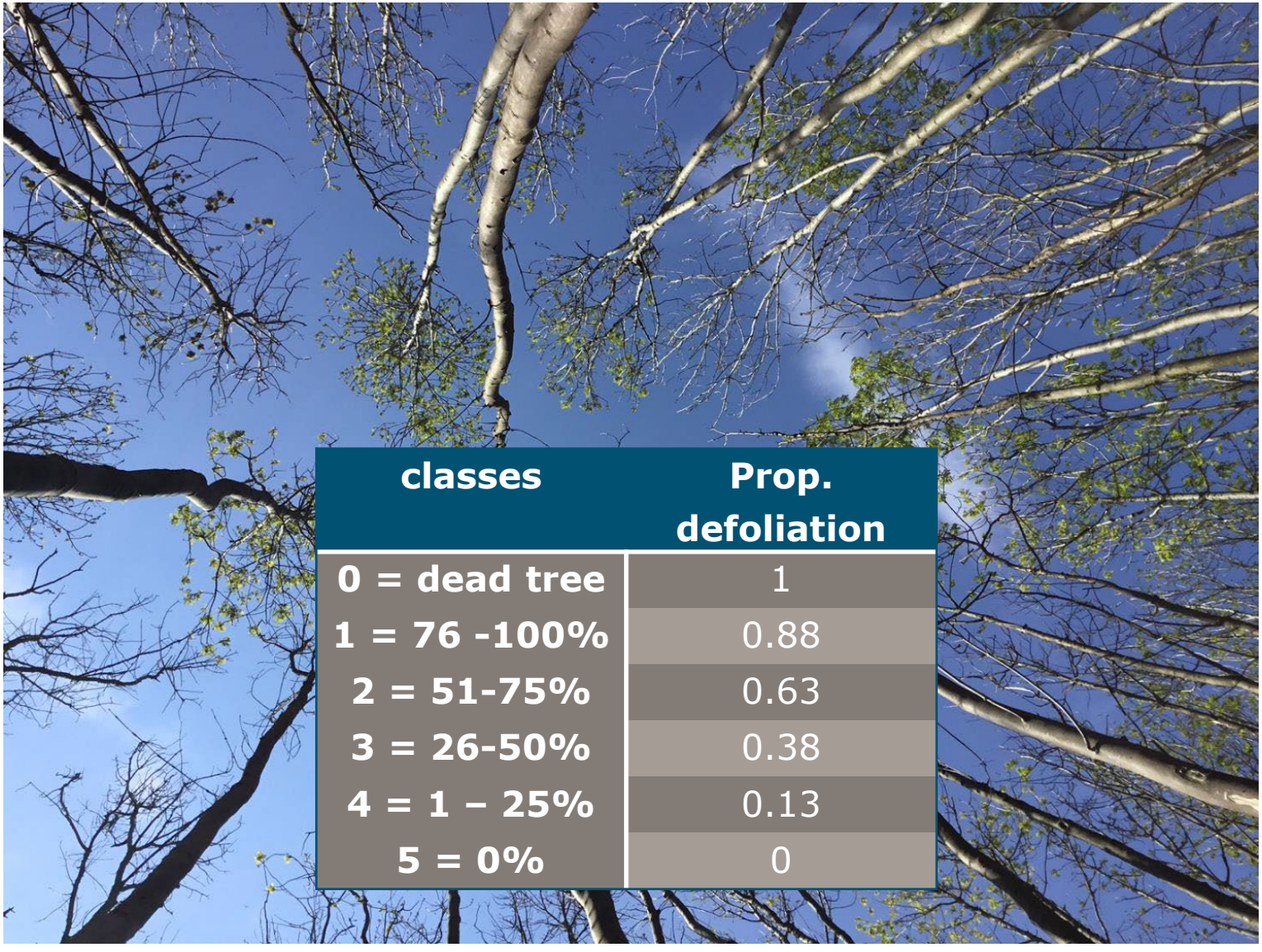
Local LD decay for chromosome 1



Constant LD decay across euchromatin.

Ash dieback: defoliation %

Based on natural infection

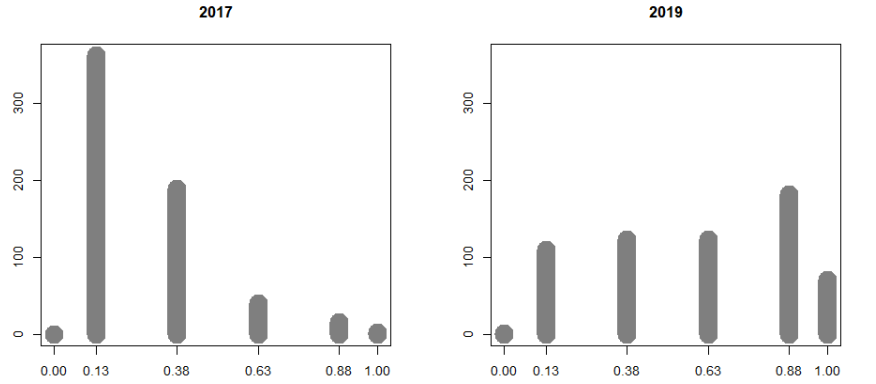


classes	Prop. defoliation
0 = dead tree	1
1 = 76 -100%	0.88
2 = 51-75%	0.63
3 = 26-50%	0.38
4 = 1 – 25%	0.13
5 = 0%	0

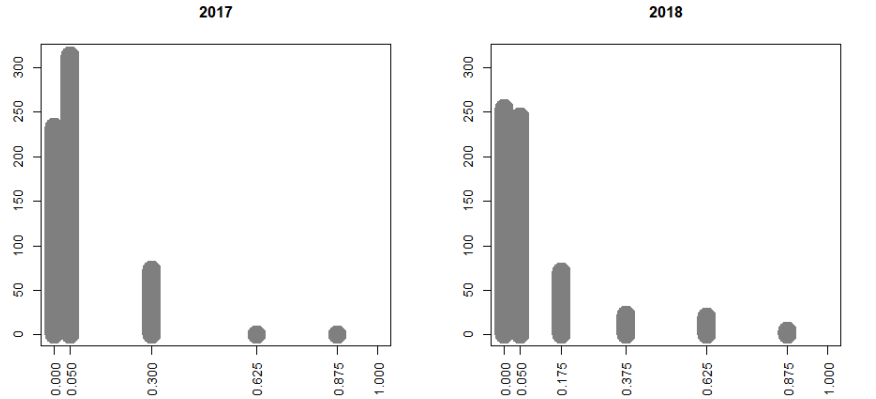
Ash dieback: defoliation %

WUR Windesheim trial (600 samples)
INRAE Moulin-sur-Yèvre trial (600 samples)

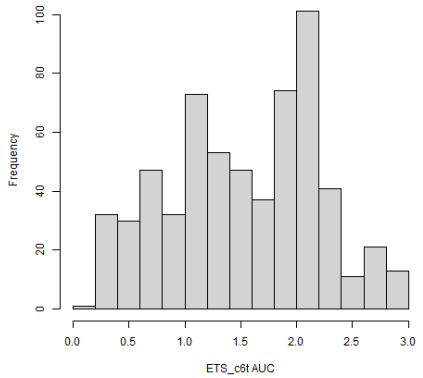
WUR

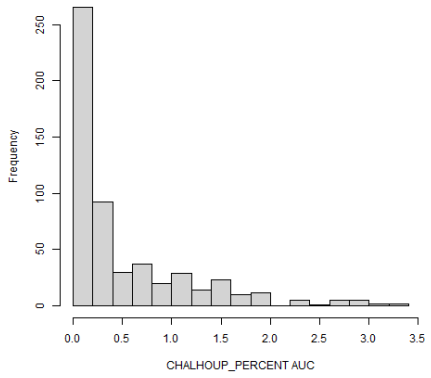


INRAE



AUC

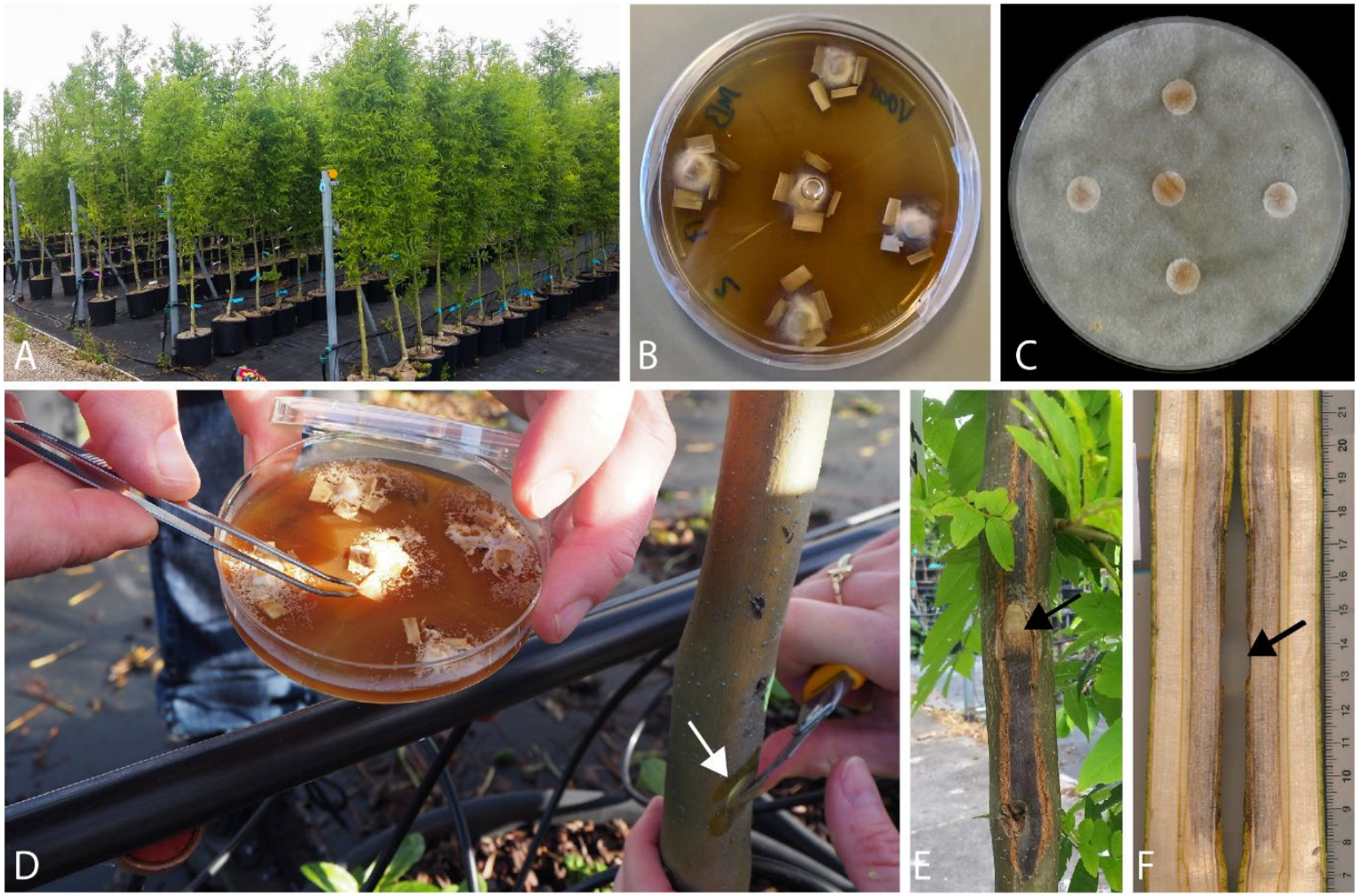




Ash dieback: lesion length

Artificial inoculation

Kowalski et al. 2015 Virulence of Hymenoscyphus albidus and H. fraxineus on Fraxinus excelsior and F. pennsylvanica. PLoS ONE 10(10): e0141592. doi:10.1371/journal.pone.0141592

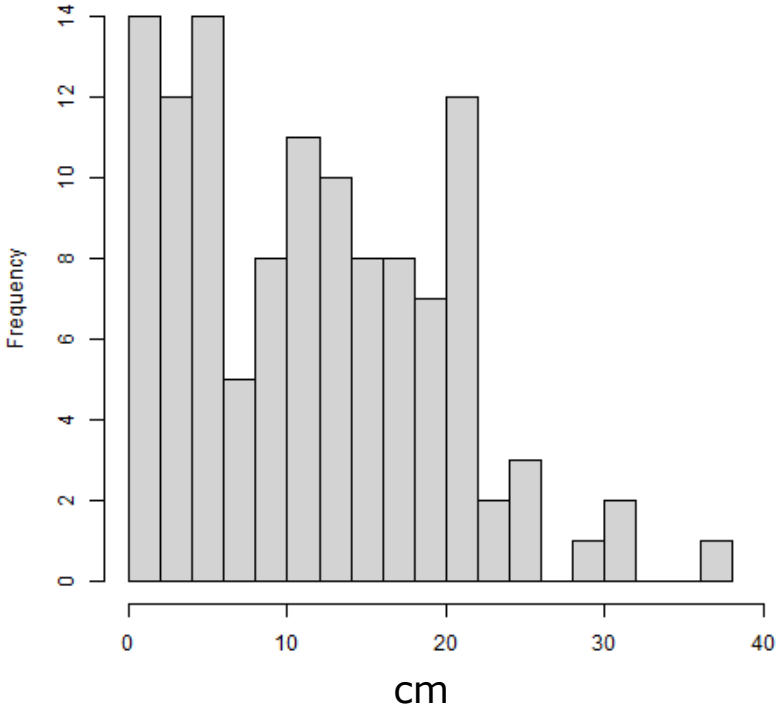


Ash dieback: lesion length

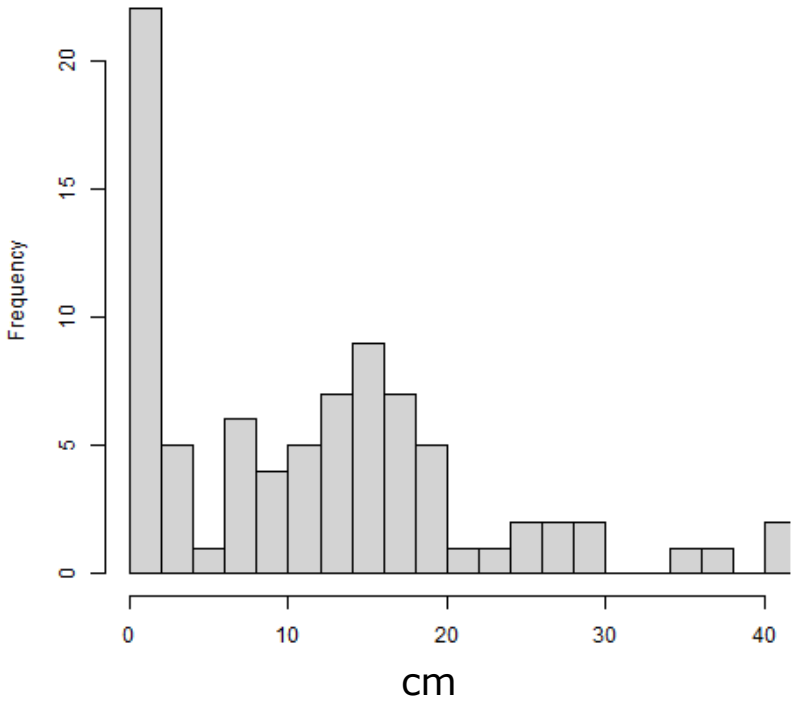
WUR 180 samples totally

Lesion length downward (cm)

2020



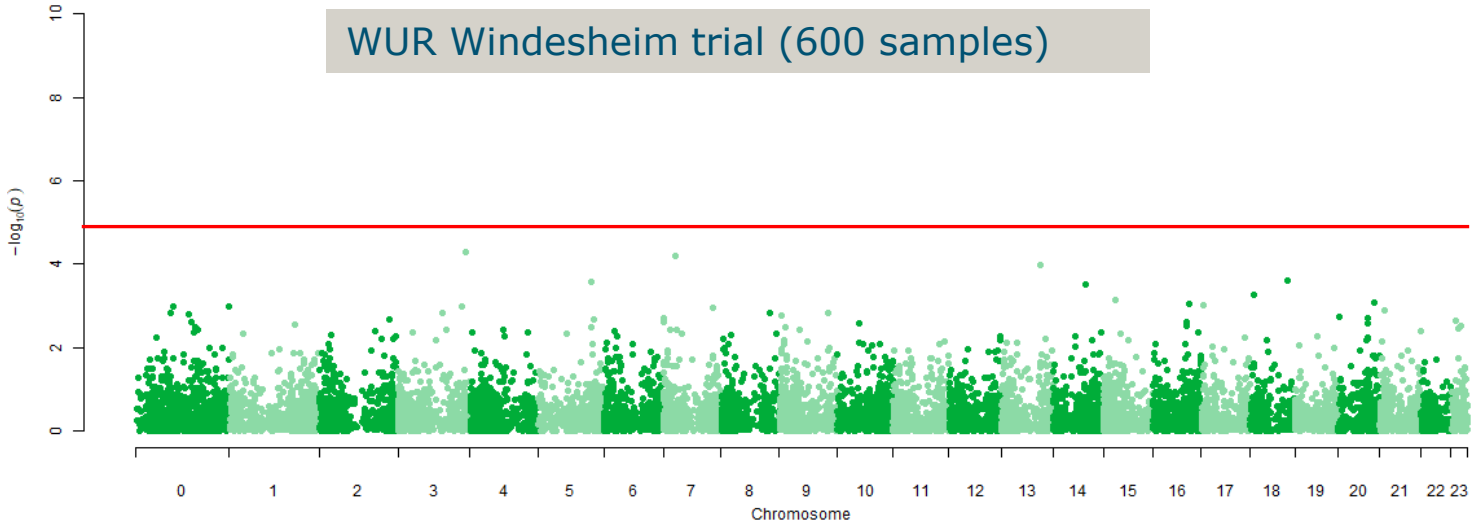
2022



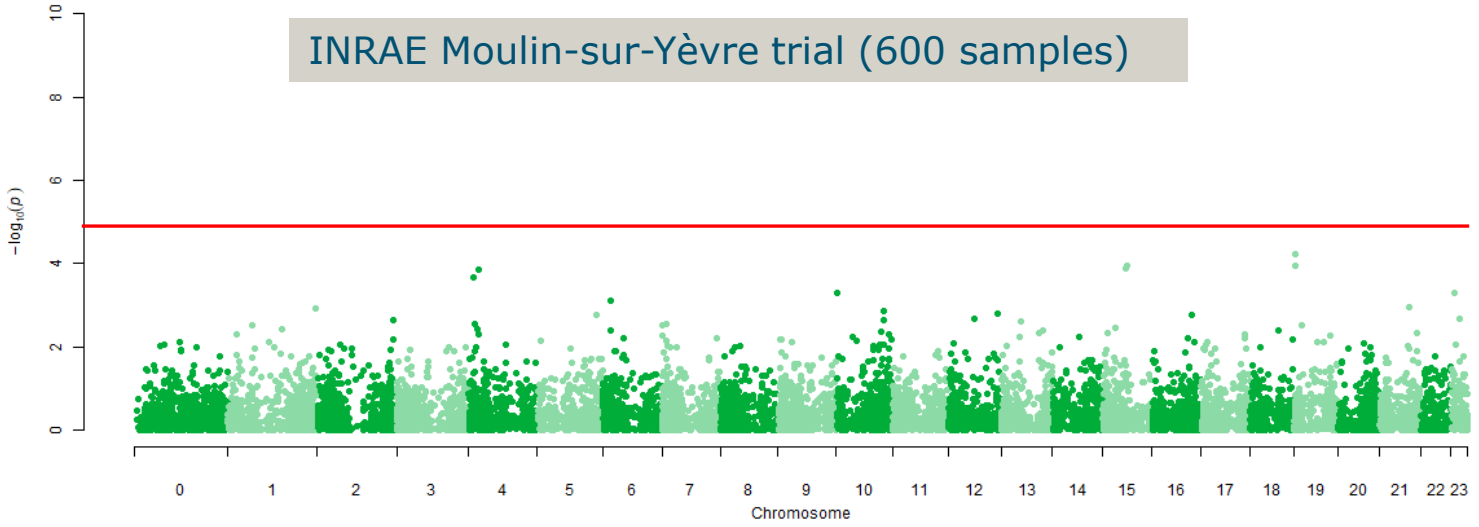
GWAS: dieback defoliation %

Single marker regression by a mixed model $y = m + \underline{K} + e$

WUR Windesheim trial (600 samples)

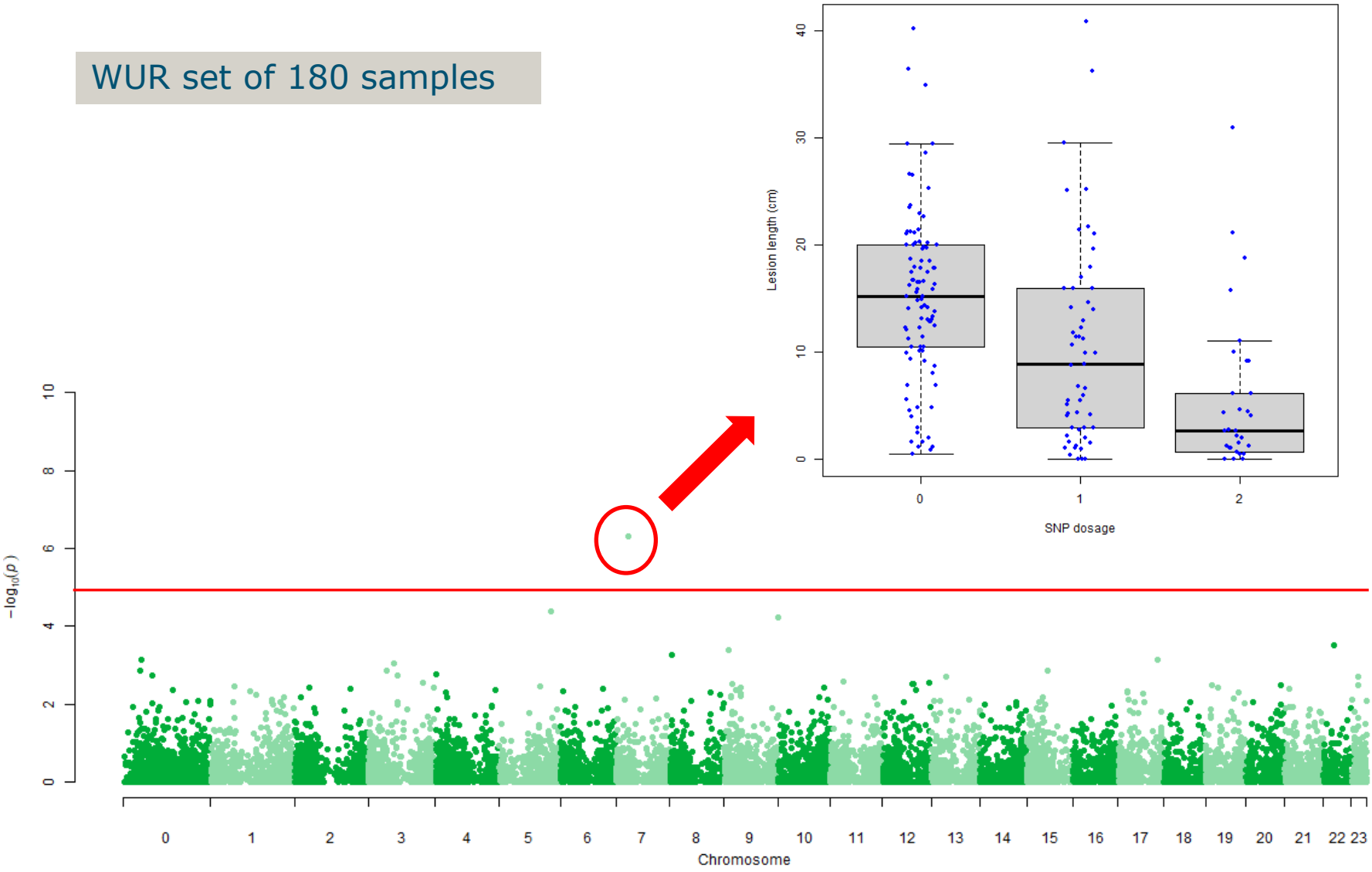


INRAE Moulin-sur-Yèvre trial (600 samples)



GWAS: dieback lesion length

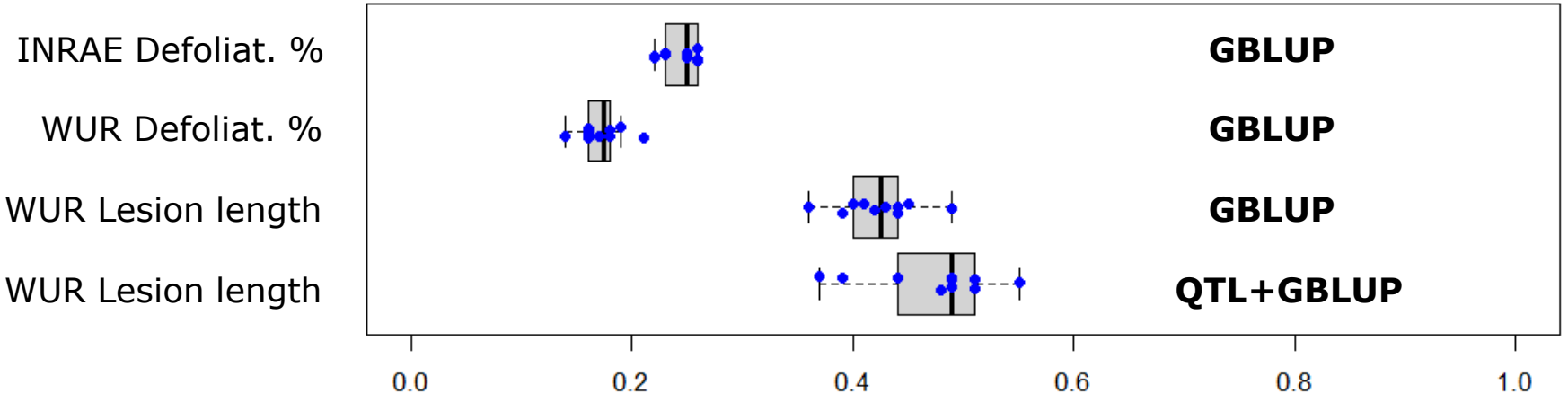
WUR set of 180 samples



Genomic prediction

Defoliation %	WUR Windesheim trial (600 samples) INRAE Moulin-sur-Yèvre trial (600 samples)
Lesion length	WUR 180 samples

Predictive abilities



10-repeats 5-fold cross-validation

Conclusions

- The 4TREE SNP array is a powerful tool for studying population structure and relatedness in ash (possible to identify full-sibs).
- Very fast LD decay (limitation for GWAS by 4TREE?).
- One significant SNP identified by GWAS for lesion length.
- Lesion length was more predictable than defoliation by using 4TREE. Although they are not alternative traits (defoliation more complex, tolerance plus escape?).
- Interesting to repeat/enlarge the artificial inoculation experiment.

Acknowledgments

Sarah Ben-Sadoun

Arnaud Dowkiw

Joukje Buiteveld

Paul Copini

Linda Kodde

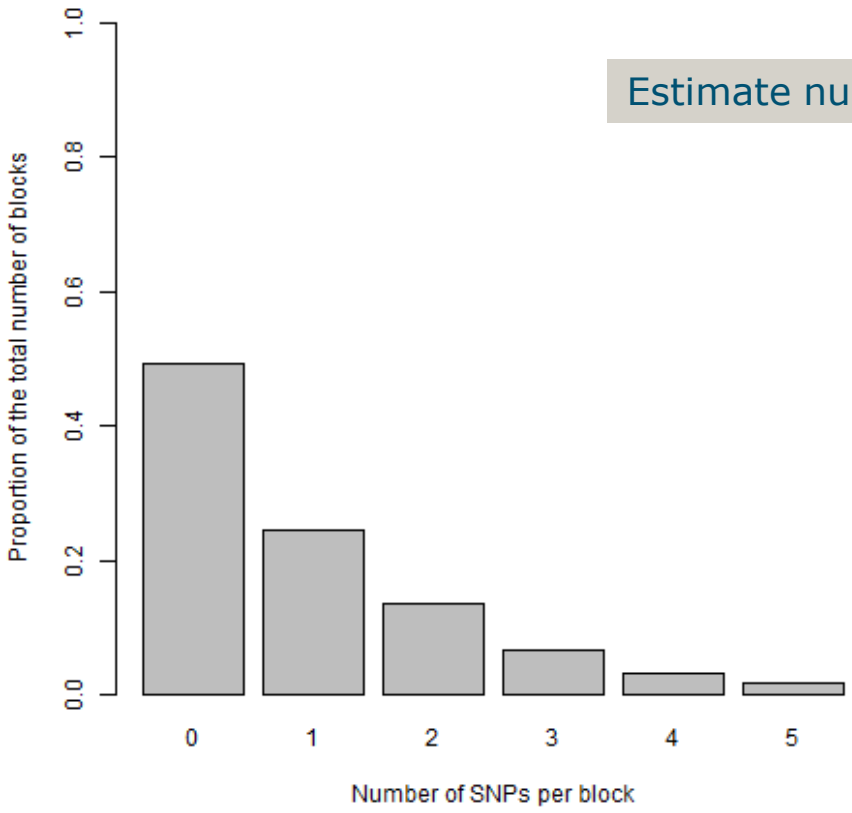
René Smulders

INRAE

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

LD decay

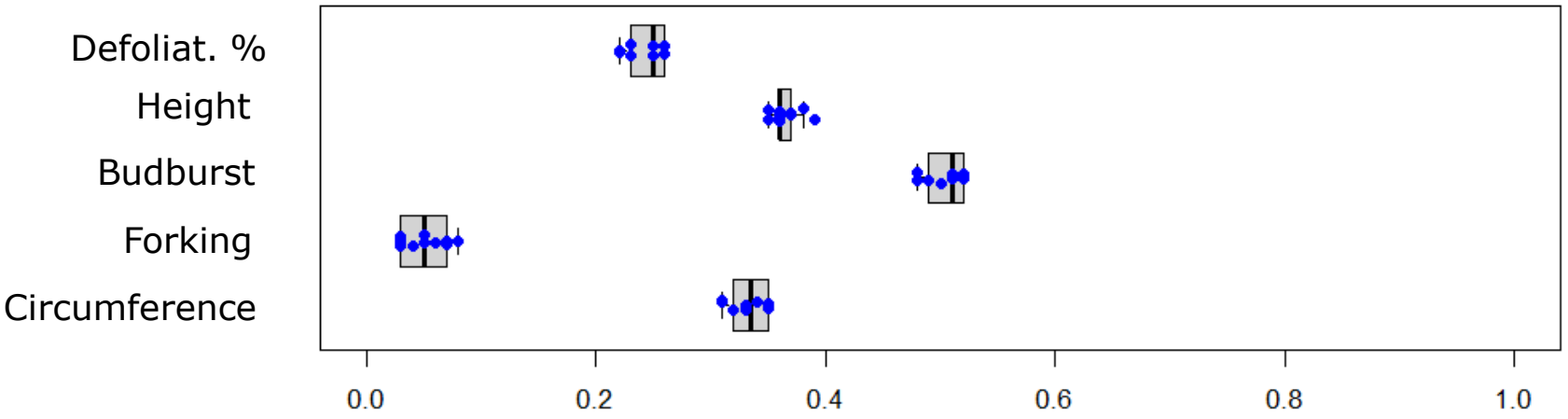
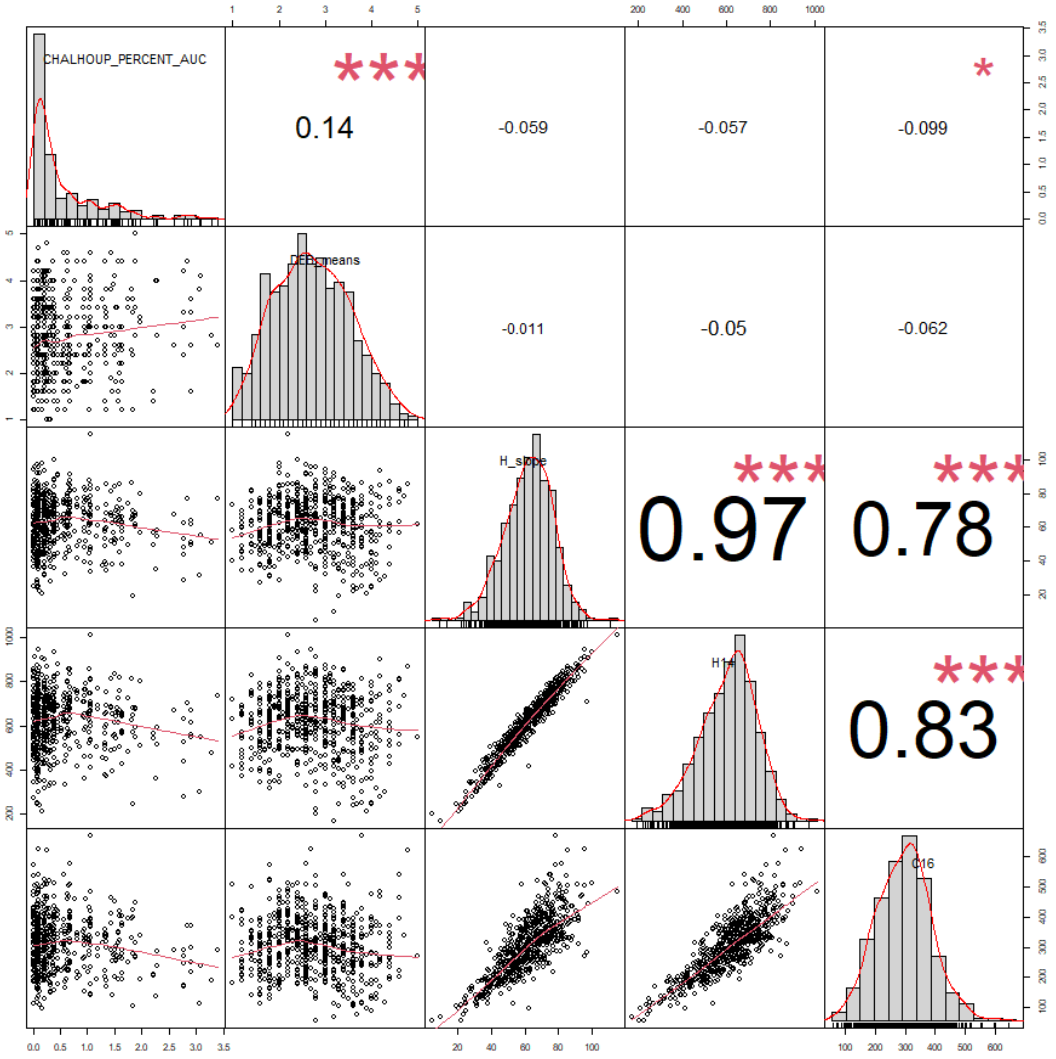


Estimate number of markers required

Genomic prediction

INRAE

Predictive abilities



Other traits (INRAE)