On the use of phylogeny-based tests to detect association between quantitative traits and haplotypes

Claire Bardel[1,3], Vincent Danjean[2], Pierre Darlu[3]
and Emmanuelle Génin[3]

[1] UMR 5145, CNRS, MNHN, Université Paris VII, Paris, France
[2] ID-IMAG, Université Joseph Fourrier, Grenoble, France
[3] INSERM U535, Villejuif, France

Introduction

Background

- Development of molecular genetics $\Rightarrow$ lots of markers available within genes
- Haplotypic methods allow to use the joint information of several markers to test for association between a gene and a disease:
  - but, increase in the number of markers $\rightarrow$ high number of haplotypes $\rightarrow$ low power of the association test
- A solution: group haplotypes
  - different grouping methods have been developed
  - in particular, the evolutionary history of haplotypes represented by a phylogenetic tree can be used

Aim of the study

Presentation of a new method to test for association between quantitative traits and haplotypes

Power study and comparison with 2 other tests by simulations

Method (1): ALTree

phenotypic values

101001... haplo 1 / 2.3 3.5 1.4

101001... haplo 2 / 0.9 1.5 3.2

011001... haplo 3 / 8.9 7.6 10.2 8.7

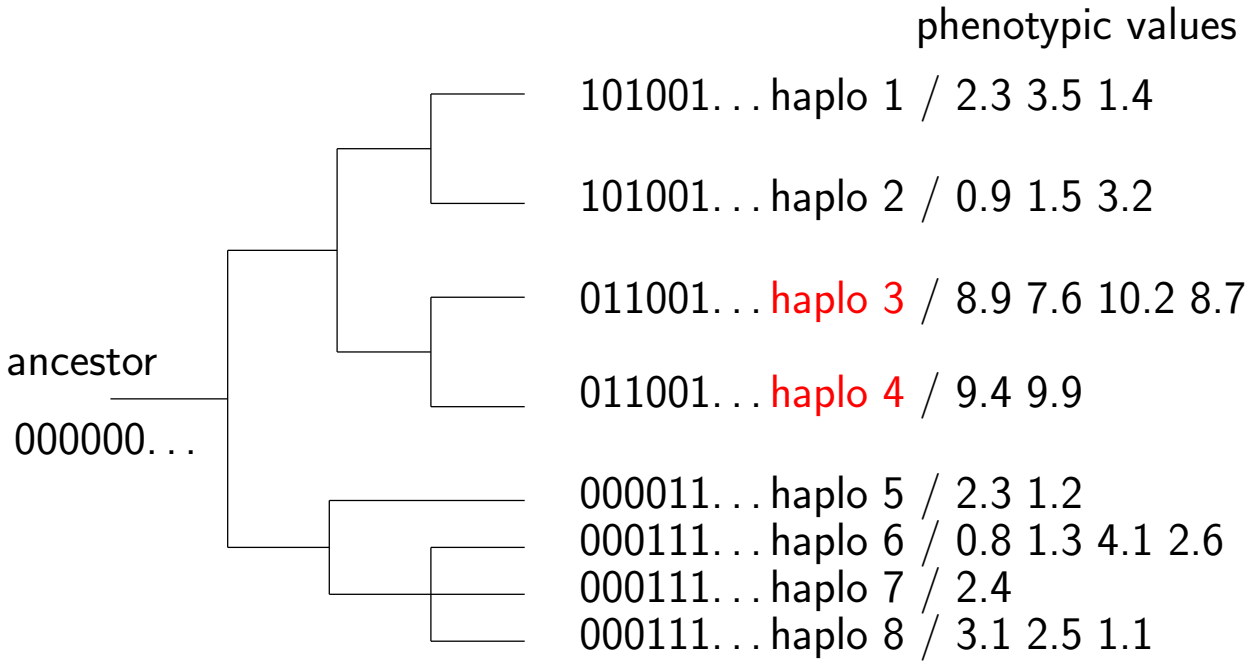
011001... haplo 4 / 9.4 9.9

000011... haplo 5 / 2.3 1.2
000111... haplo 6 / 0.8 1.3 4.1 2.6
000111... haplo 7 / 2.4
000111... haplo 8 / 3.1 2.5 1.1

The data

- Haplotypes formed by SNPs or microsatellites
- The quantitative trait value of each individual is assigned to his two haplotypes

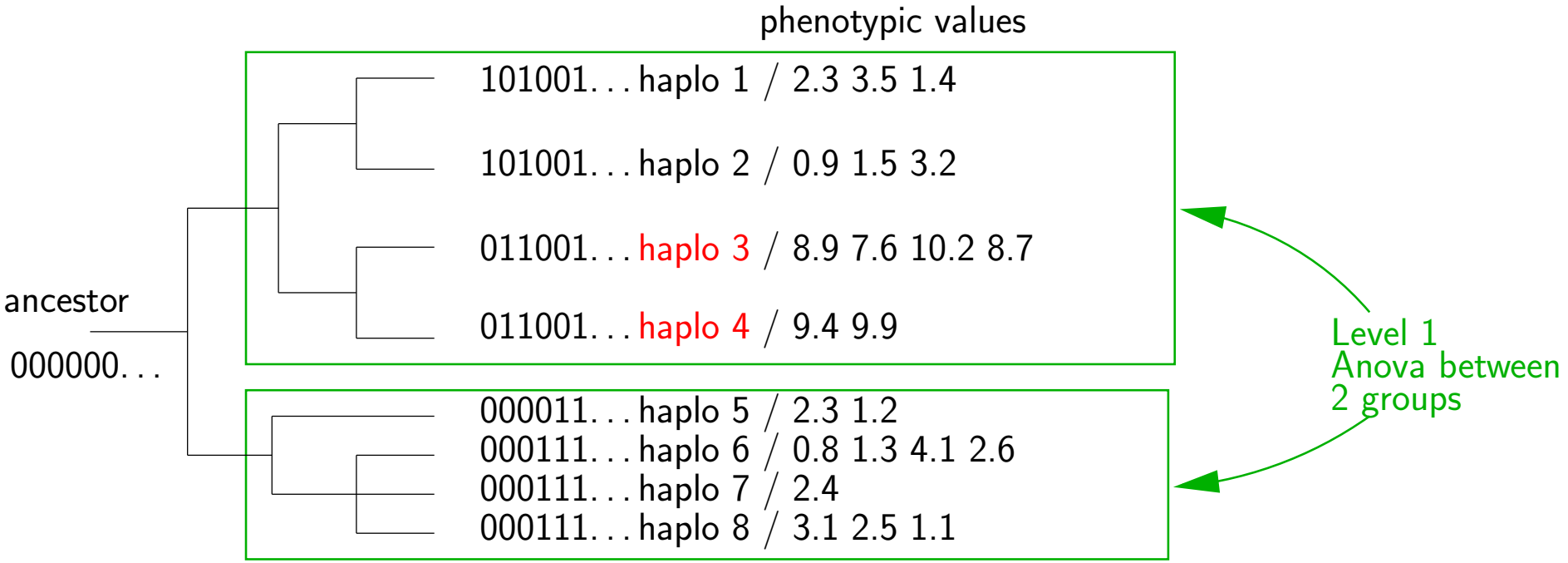
Method (1): ALTree



Reconstruction of the phylogenetic tree

- Use of a parsimony method (software PAUP*) or a ML method (software PHYML)
- Rooting of the tree
  - here, the most frequent haplotype is used as the ancestral sequence

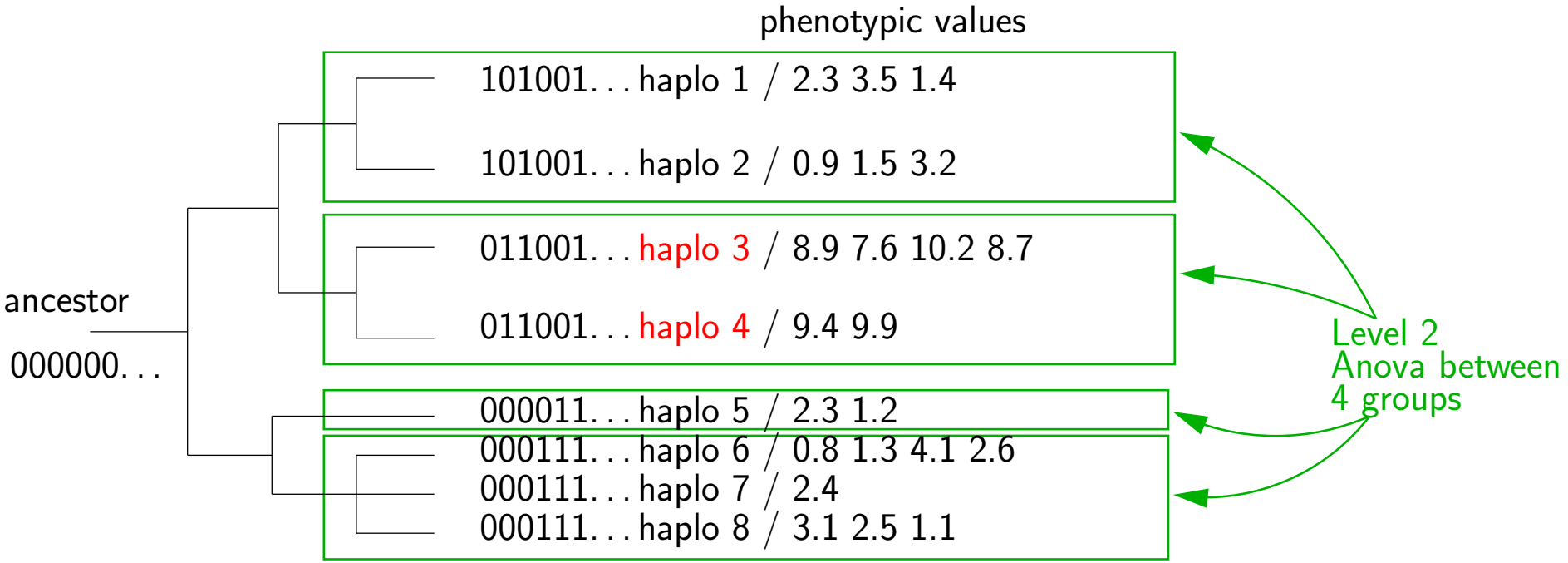
Method (1): ALTree



Nested analysis of the tree

- Analysis starts from the root
- At each level: a one-way ANOVA between the different groups defined on the tree is performed
- Correction for multiple testing using the permutation procedure of Ge *et al*, 2003 $\longrightarrow$ One p-value for a tree

Method (1): ALTree



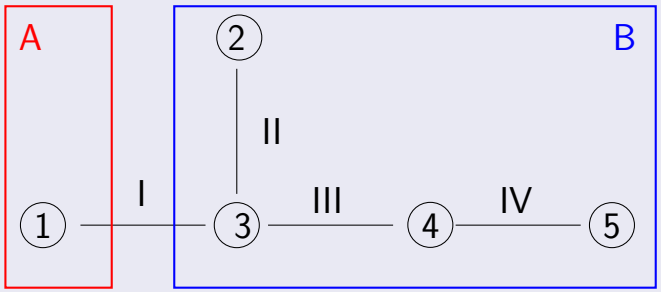
Nested analysis of the tree

- Analysis starts from the root
- At each level: a one-way ANOVA between the different groups defined on the tree is performed
- Correction for multiple testing using the permutation procedure of Ge *et al*, 2003 $\longrightarrow$ One p-value for a tree

Method (2): The 2 other tests

TreeScan (Templeton et al, 2005)

- It works on an unrooted tree
- For each partition of the tree defining two groups A and B:

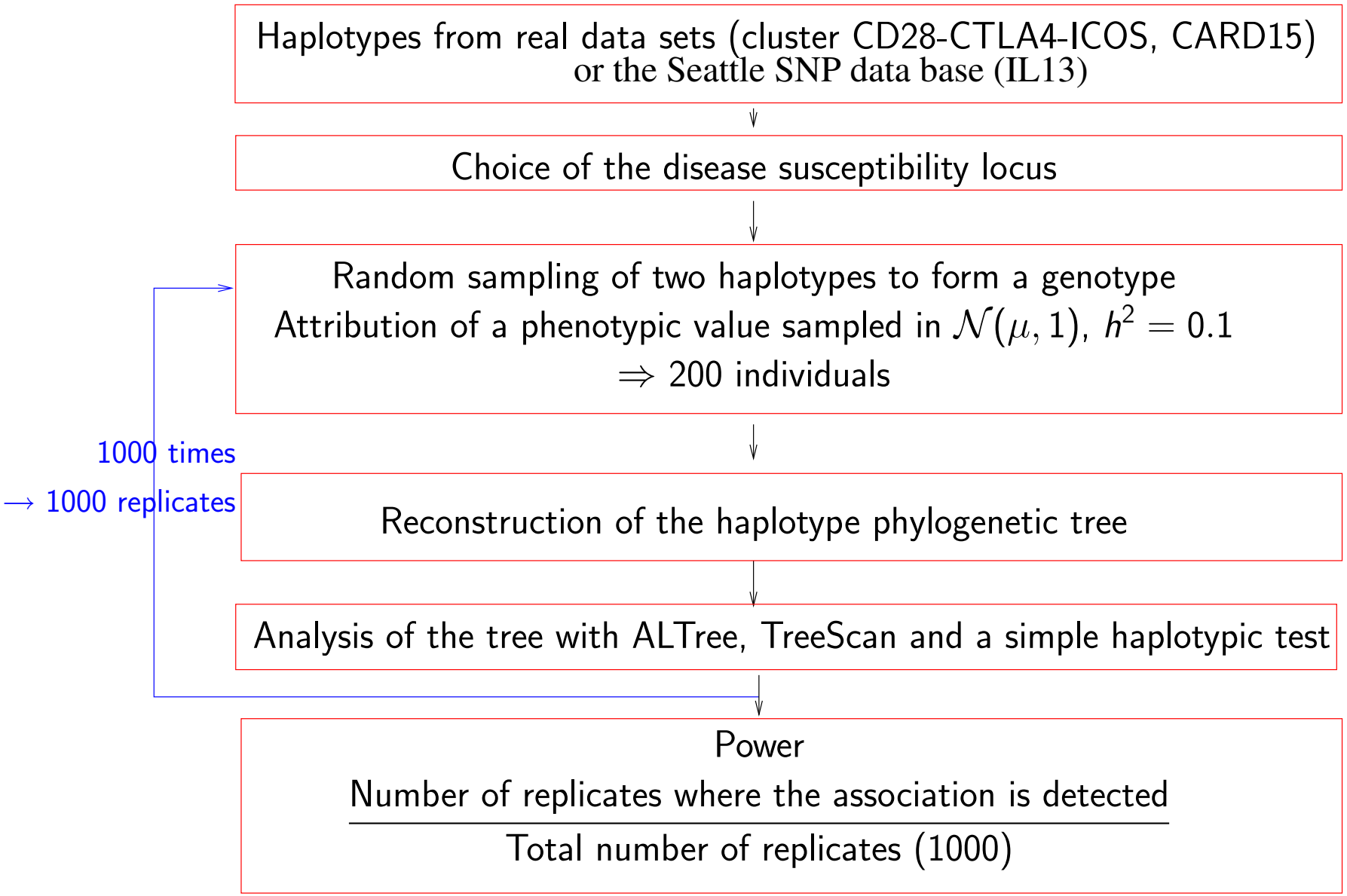


  - 1/ Put each individual into one of the three groups: AA, AB and BB depending on his haplotypes
    - ex: H1H5 -> group AB
  - 2/ Perform an ANOVA between the quantitative values in the three groups
  - 3/ Compute the p-value by permutations
- Then, use the correction for multiple testing of Westfall and Young (1993)

An omnibus haplotypic test

- One-way ANOVA between all the haplotypes
- p-value determined by permutations

The simulation process



Result (1): power

Power of the 3 methods in the 3 regions

- Additive model, DS site removed, $h^2 = 0.1$, 200 individuals
- Average power (%) over each SNP in turn considered as the DS site

	# SNPs	ALTree	TreeScan	Haplo
CARD15	13	57.6	46.5	51.7
CTLA4 clus.	17	59.8	44.3	57.3
IL13	12	77.9	59.0	75.6
mean		65.1	49.9	61.5

Conclusion

- For the 3 methods: higher power for the IL13 gene
- In these conditions, ALTree is the most powerful test
- Type I errors of the three tests are between 4.3% and 5.6%

Results (2): influence of the genetic model

Comparison of the 3 methods for 3 models, $h^2 = 0.1$

times a method is more powerful than the two others:

	ALTree	TreeScan	Haplo
DS site removed			
Dominant	46%	17%	37%
Additive	51%	5%	44%
Recessive	17.5%	47.5%	35%
DS site present			
Additive	59%	17%	24%

- Analysis of the 42 loci
- h^2=0.1
- 200 indivduals

Conclusion

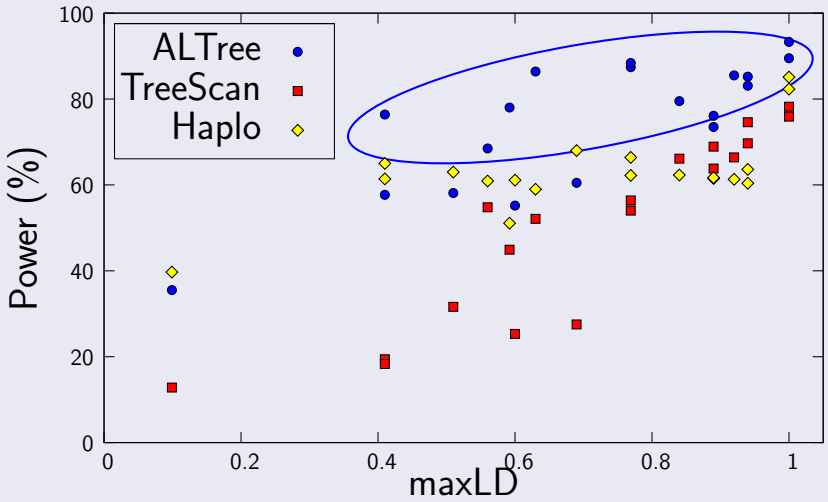
- ALTree: more powerful for dominant and additive models
- TreeScan: more powerful for recessive models
- DS site present: power of the phylogeny-based methods is increased

Results (3): Influence of maxLD and DS allele frequency

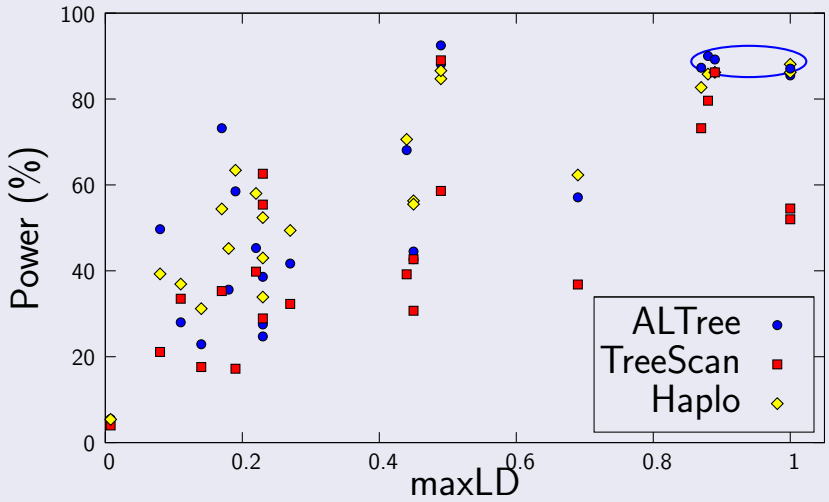
Power of the 3 methods according to the maxLD

- Simulations with the DS site removed, additive model
- maxLD: highest value of LD observed with the DS site

High DS allele frequency (>.25)



Low DS allele frequency (<.25)



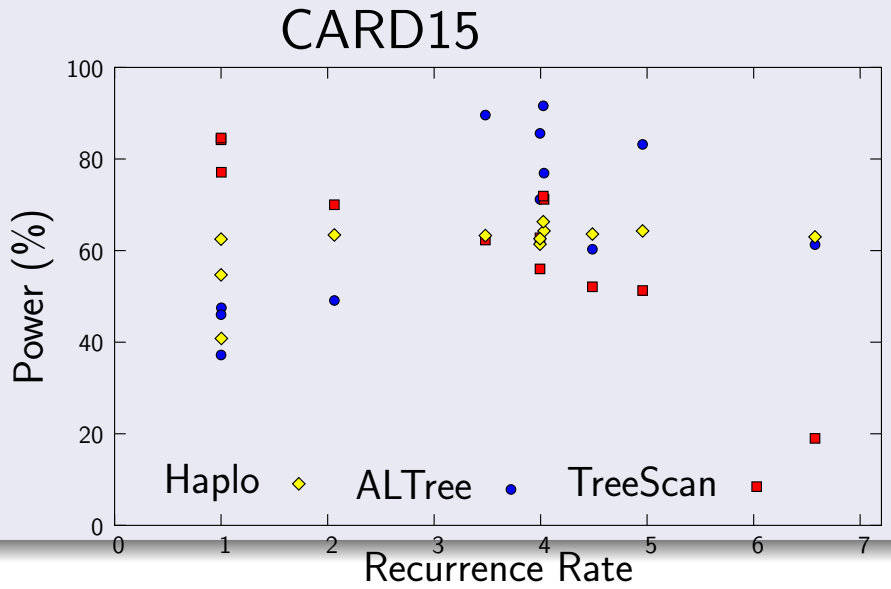
Conclusion

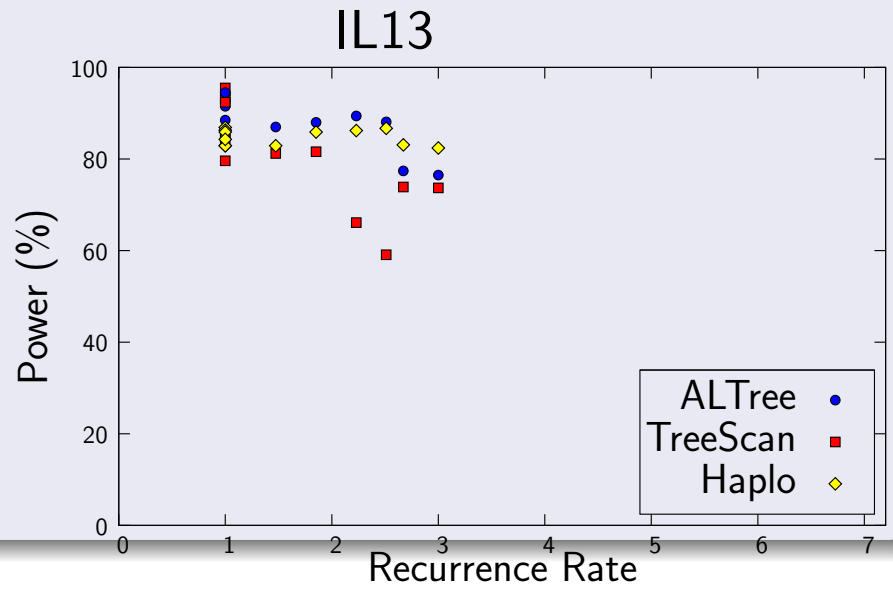
- As expected: increase in the power of the 3 methods with the maxLD
- ALTree more powerful when the frequency of the DS allele is high
- ALTree more powerful when the maxLD is high

Results (4):Influence of the recurrence rate of the DS site

Power of the 3 methods according to the recurrence rate

- Simulation with the DS site, Additive model
- Recurrence rate: # of times the DS locus mutates in the tree



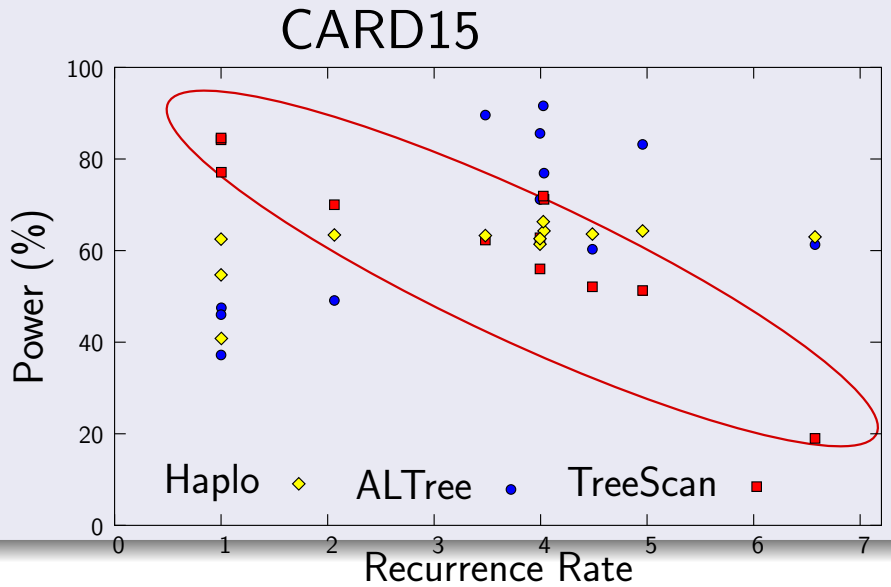


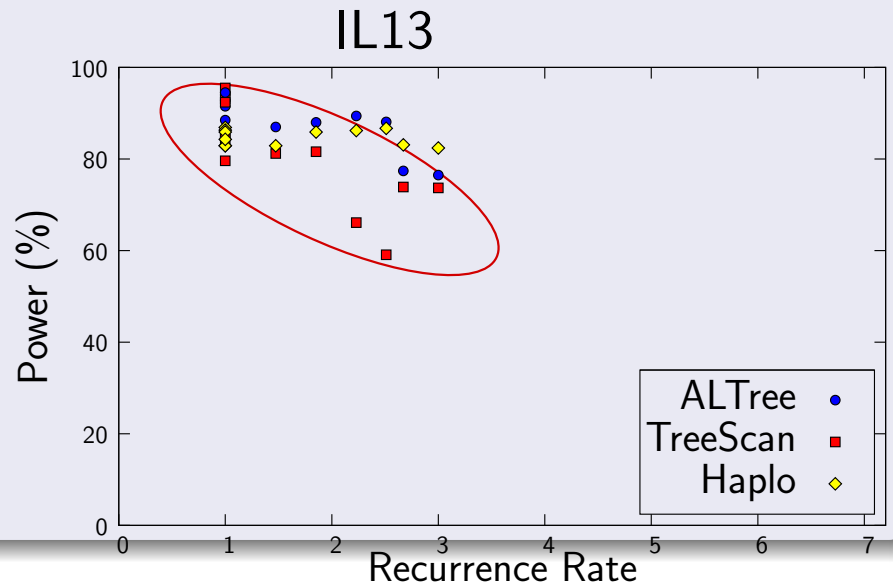
Conclusion

Results (4):Influence of the recurrence rate of the DS site

Power of the 3 methods according to the recurrence rate

- Simulation with the DS site, Additive model
- Recurrence rate: # of times the DS locus mutates in the tree





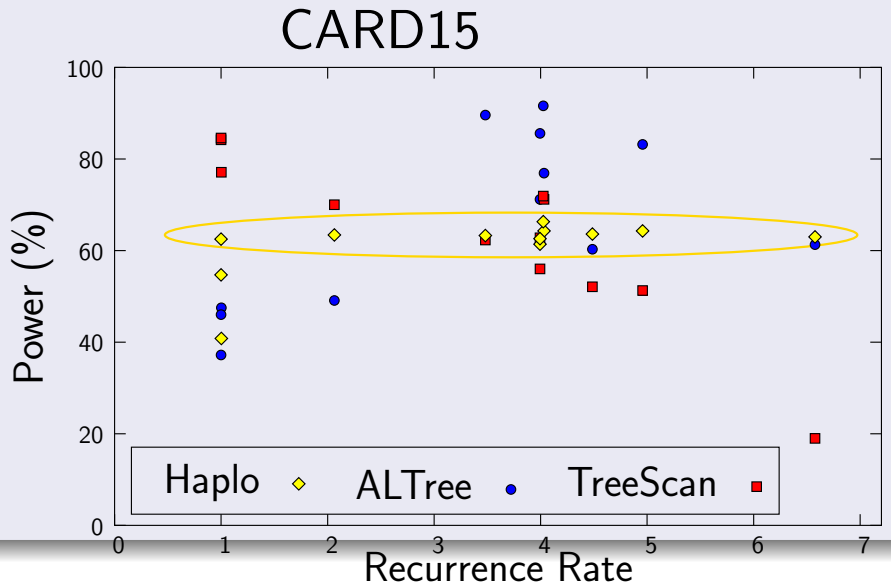
Conclusion

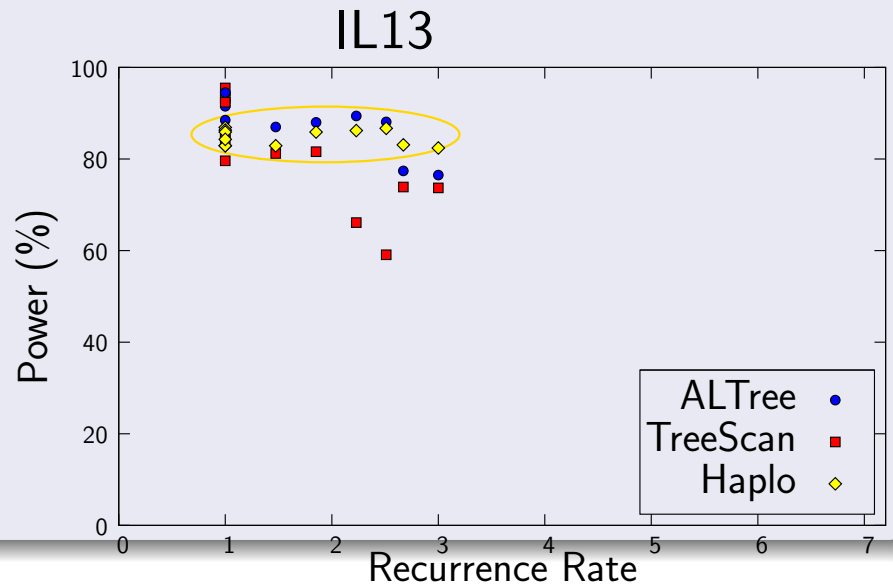
- TreeScan: decrease in power when the recurrence rate increases

Results (4):Influence of the recurrence rate of the DS site

Power of the 3 methods according to the recurrence rate

- Simulation with the DS site, Additive model
- Recurrence rate: # of times the DS locus mutates in the tree





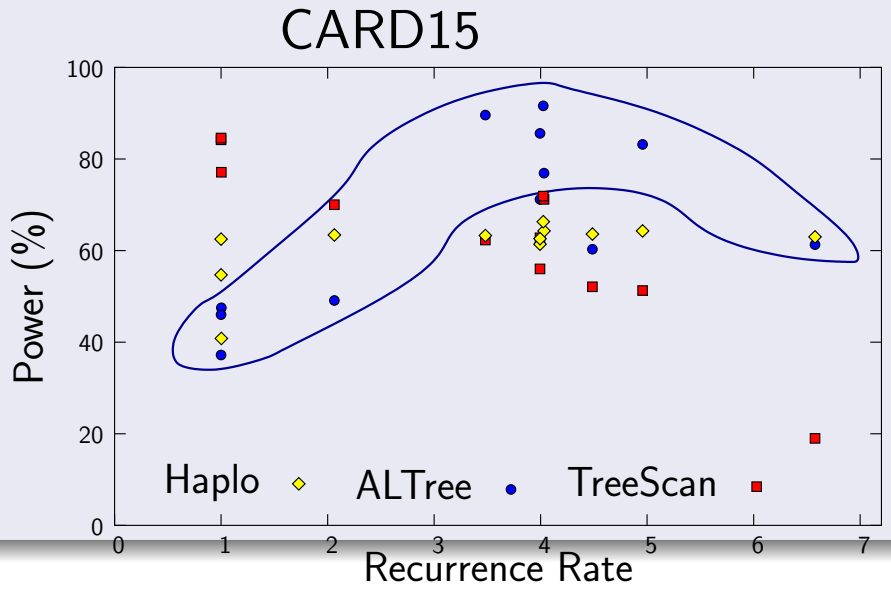
Conclusion

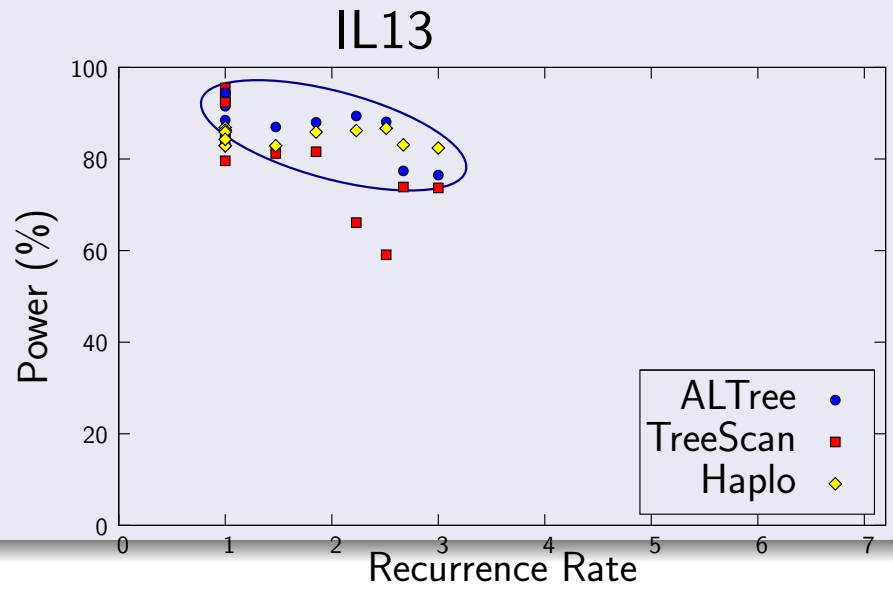
- TreeScan: decrease in power when the recurrence rate increases
- Haplo: No influence of the recurrence rate

Results (4):Influence of the recurrence rate of the DS site

Power of the 3 methods according to the recurrence rate

- Simulation with the DS site, Additive model
- Recurrence rate: # of times the DS locus mutates in the tree





Conclusion

- TreeScan: decrease in power when the recurrence rate increases
- Haplo: No influence of the recurrence rate
- ALTree: Influence varies with the studied gene

Conclusion

- **Development** of a new association test
  Implementation in our software ALTree (quantitative test available soon) `http://Claire.Bardel.free.fr/software`
- **Study of its power**: it increases with
  - the sample size
  - the heritability of the trait (not shown)
  - the DS allele frequency
  - the linkage disequilibrium
- **Comparison** with TreeScan and the omnibus haplotypic test
  - ALTree is more powerful for additive and dominant models
  - ALTree is more powerful for high DS allele frequency
  - DS removed: ALTree is more powerful when the LD with the DS site is high
  - DS present: ALTree is less dependant on the recurrence rate of the mutations than TreeScan

Future work

Methodological work

- Development of a new test based on hierarchical ANOVA
  - Study its power and compare it with the one way ANOVA, TreeScan and the haplotypic test
- Study the power of these methods when the trait is influenced by the interaction of several loci
- Study the impact of the haplotype reconstruction on the method and possibly, take into account the uncertainty in haplotype reconstruction

Application to data sets

- Replication of a study on TAFI (Thrombin Activatable Fibrinolysis Inhibitor)
  - first results: detection of the association and identification of loci reported to be involved in the determinism of the quantitative trait