The use of markers and perspectives for genomic selection in maritime pine

Laurent Bouffier

The use of markers and perspectives for genomic selection in maritime pine (*Pinus pinaster*)

Laurent Bouffier
INRA (French National Institute for Agricultural Research)
laurent.bouffier@inra.fr

Forestry Workshop on Genomic Selection - Implementation and Impacts for the UK
17 September 2019 - Edinburgh

Outlines

1/ French maritime pine breeding programme

2/ Molecular markers available

3/ Identity and pedigree checking

4/ Mating designs based on pedigree recovery

5/ GS proof-of-concept in maritime pine

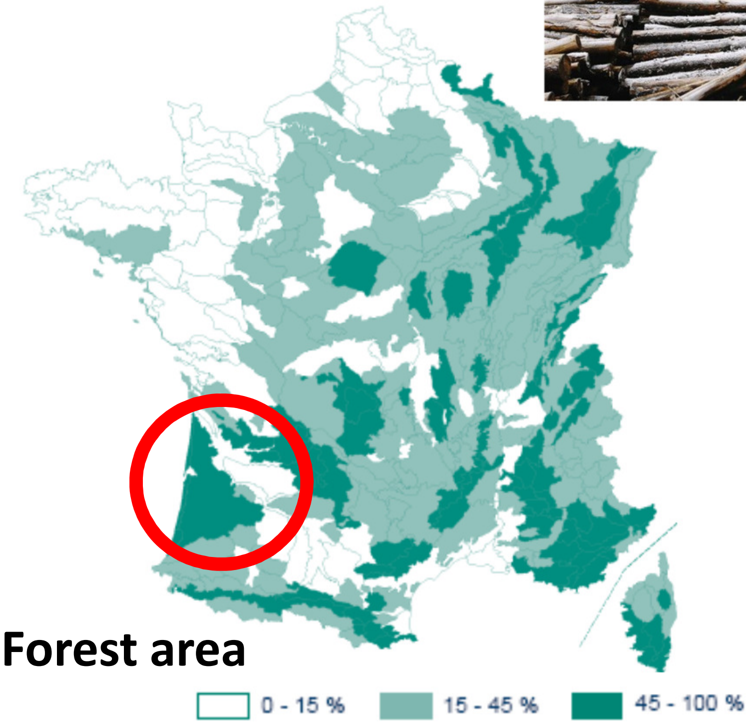
6/ Perspectives for GS in maritime pine

1/ French maritime pine breeding programme



Landscapes in southwestern France

French vineyards

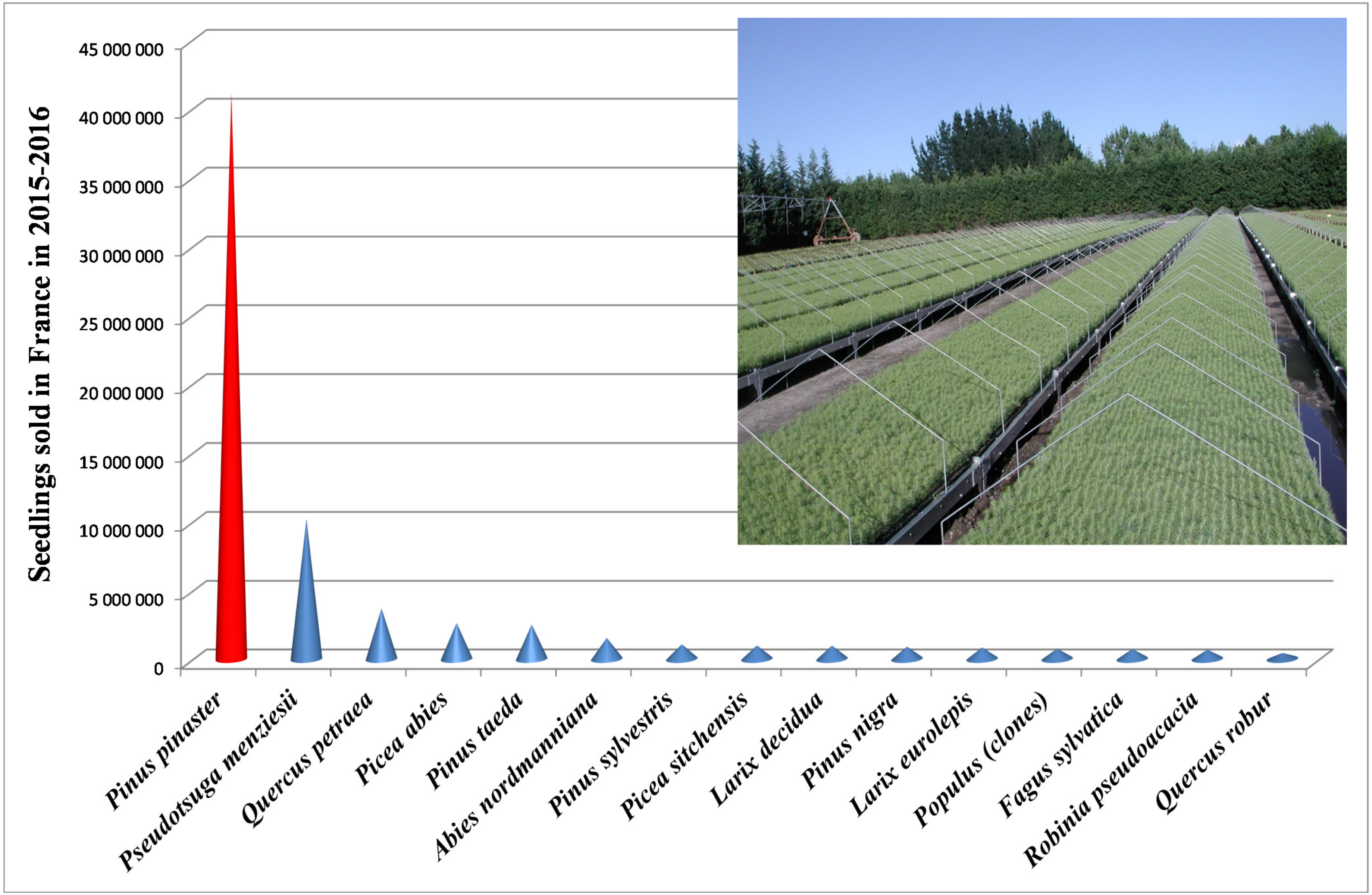


Forest area

Maritime pine forest in southwestern France

- **Maritime pine forest established at the end of the 19th century**
  - Native species
  - Poor soils (sandy podzol)
  - Dry summers / wet winters (hydromorphic soils)
- **0.8 million hectares forest (7% of the French forest)**
  - Maritime pine plantations (>90% improved seedlings) except in sand dunes (natural regeneration)
- **24% of French wood harvest**
  - 11 m3/ha/year in average
  - 8.5 million m^3 harvested /year
  - 60% saw timber / 40% industrial wood

Maritime pine = 1st plantation species in France



Breeding program managed by a consortium

- Base population established in 1960’s by INRA
- Two breeding programs (INRA, FCBA)
- Consortium since 1995 = “ Groupe Pin Maritime du Futur ”

« Recurrent » fundings = Ministry of Agriculture + Région « Nouvelle Aquitaine »
Specific studies = European Union + Ministry of Education and Research (ANR projects)

Stakeholders for maritime pine breeding

Maritime pine breeding program

Managed by GPMF (« Groupe Pin Maritime du Futur »)

IP = INRA + FCBA

Seed Orchards

Composition + Design

GPMF but SO must be validated by a national public authority (CTPS)

Ownership + Management

Public (ONF) and private companies

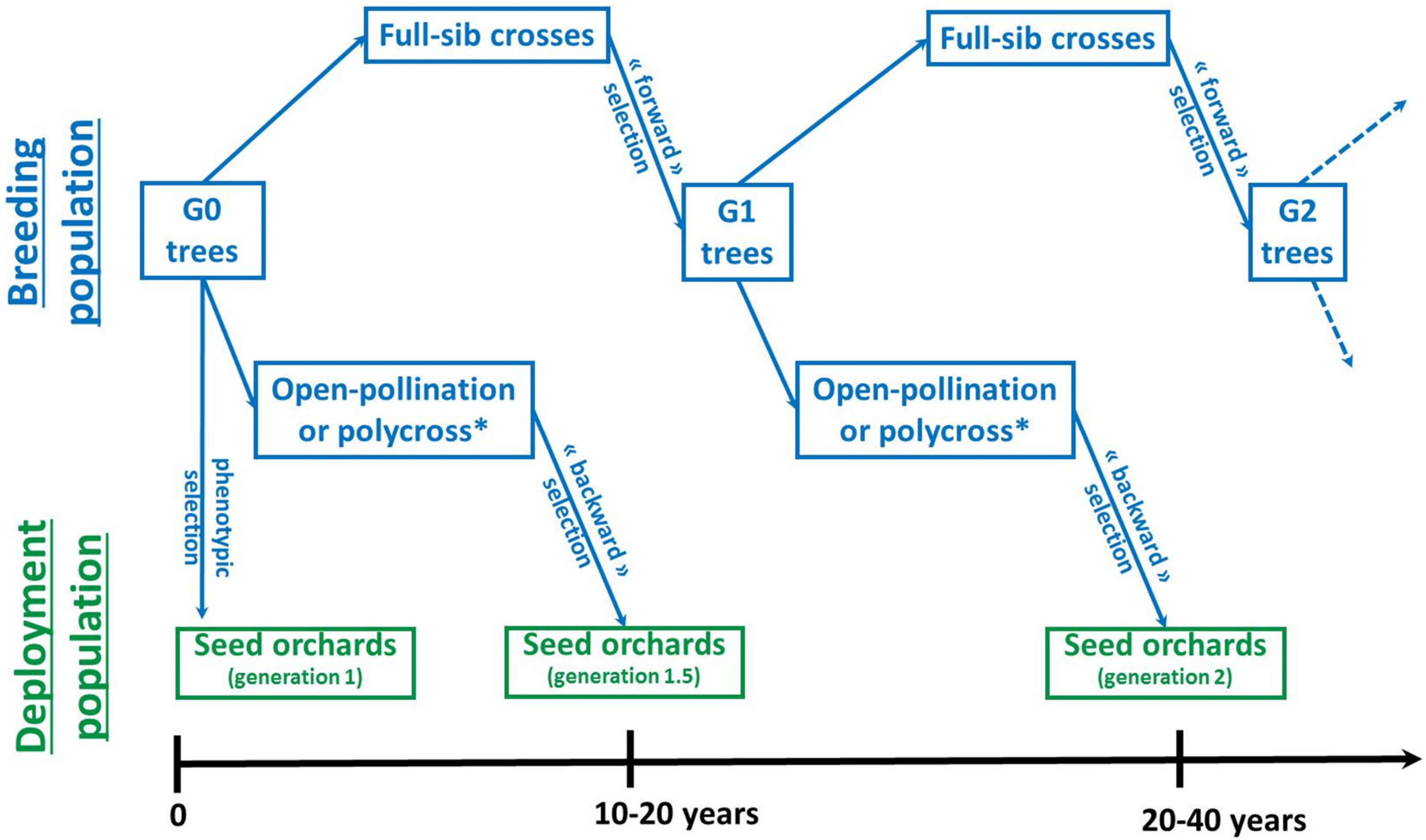
royalties

Nurseries

Private companies

(Forelite, Planfor, Naudet, Robin)

Recurrent selection scheme



- **Selection criteria:**
  - Volume (girth, height)
  - Stem straightness
  - Rust resistance
  - Wood quality traits (branching quality, wood density, spiral grain)

Maritime pine breeding: the latest advances

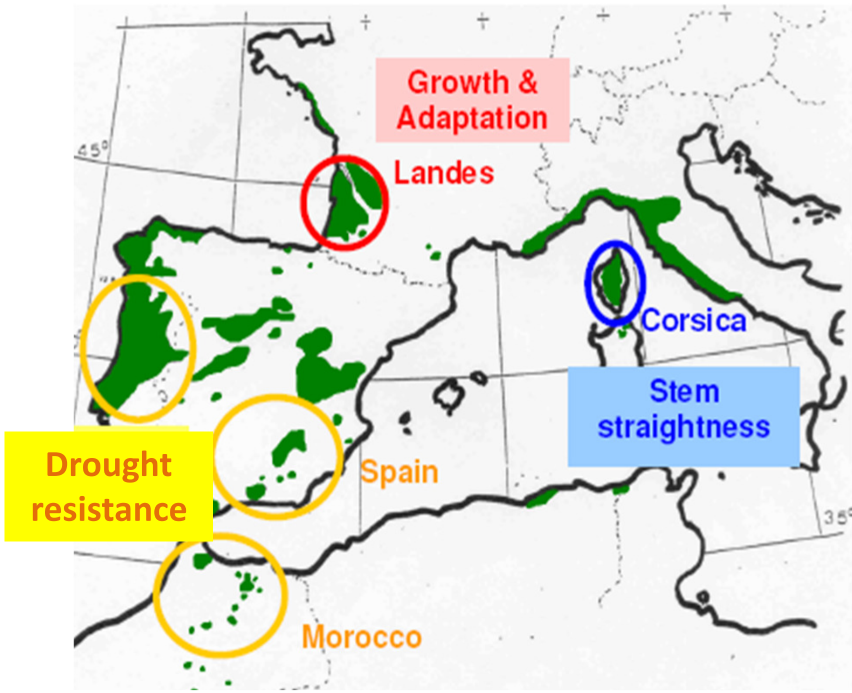
- **BLUP evaluation** (individual mixed model based on pedigrees and phenotypes)
  - database = 600,000 trees including historical trials (first one in 1962)
  - powerful genetic evaluation

- **Seed orchards renewed more often**
  - previously: one new SO composition every 15 years based on backward selections
  - currently: faster turnover (~3 years) based on backward and forward selections

- **New selection criteria under study**
  - pest resistance (pine wood nematode)
  - climate change (drought resistance)

- **New provenances**
  - Genetic diversity infusion with southern provenances for drought resistance

- **Rolling front strategy**



2/ Molecular markers available

High density genotyping

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 574–587 doi: 10.1111/1755-0998.12464

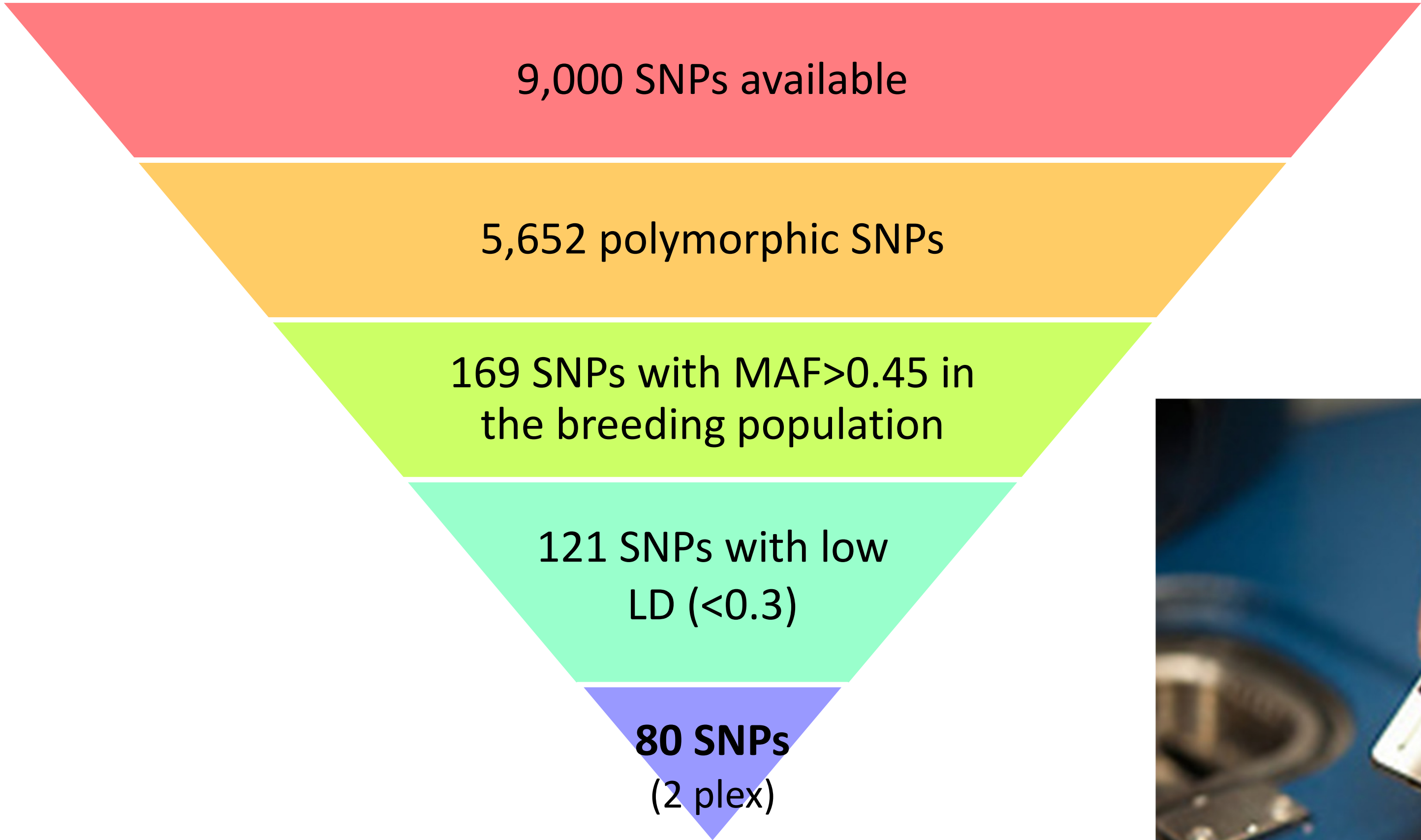
High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*)

C. PLOMION,*† J. BARTHOLOMÉ,*† I. LESUR,*‡ C. BOURY,*† I. RODRÍGUEZ-QUILÓN,§ H. LAGRAULET,*† F. EHRENMANN,*† L. BOUFFIER,*† J. M. GION,*¶ D. GRIVET,§ M. DE MIGUEL,*† N. DE MARÍA,§ M. T. CERVERA,§ F. BAGNOLI,** F. ISIK,†† G. G. VENDRAMIN** and S. C. GONZÁLEZ-MARTÍNEZ§

- Currently: **9,000** Illumina Infinium SNP array (transcriptome data + candidate genes)

- In 2020: **12,500** SNPs available (B4EST project) with a multi-species commercial array (~50€/sample)

Low density genotyping (80 SNPs)



Genotyping using Sequenom's Mass Array technology (mass spectroscopy)

(cost ~7€ per sample)

Applications High / Low density genotyping

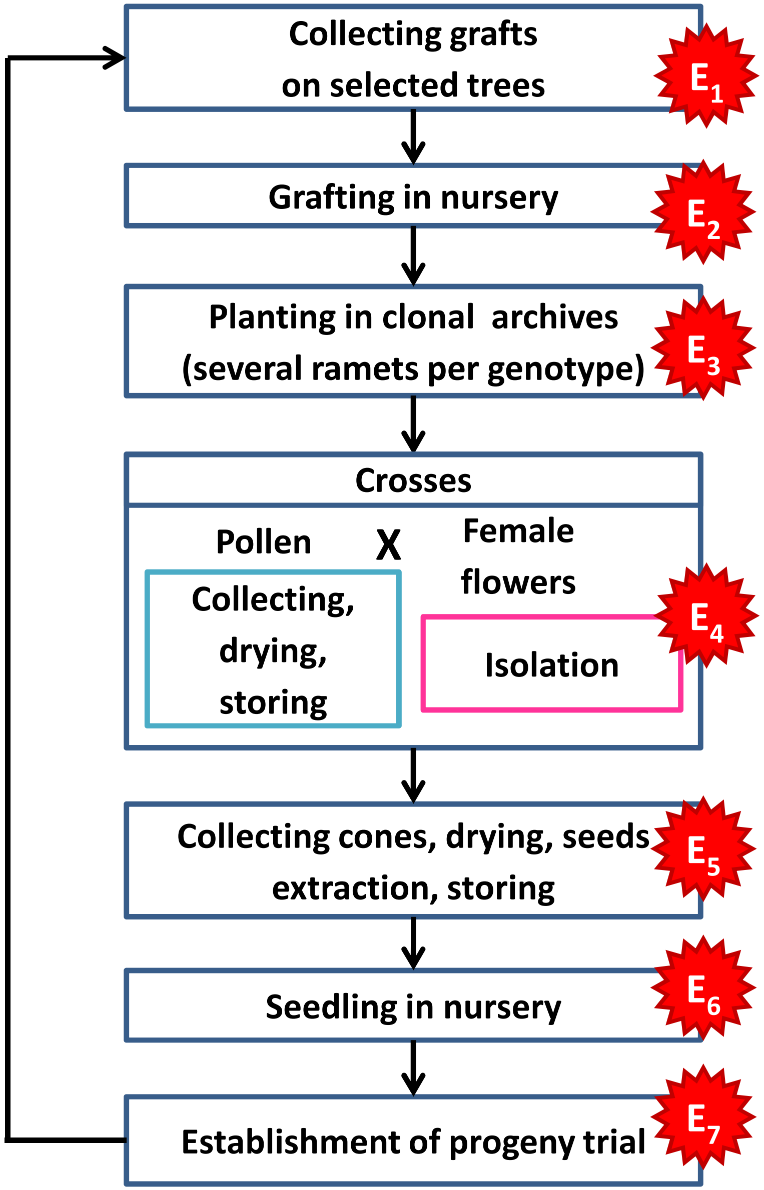
- **High density genotyping for:**
  - linkage mapping
  - association studies
  - genomic selection

- **Low density genotyping for:**
  - id checking
  - pedigree recovery
  - pollen contamination / parental contribution in SO
  - seedlot certification (provenance)

3/ Identity and pedigree checking

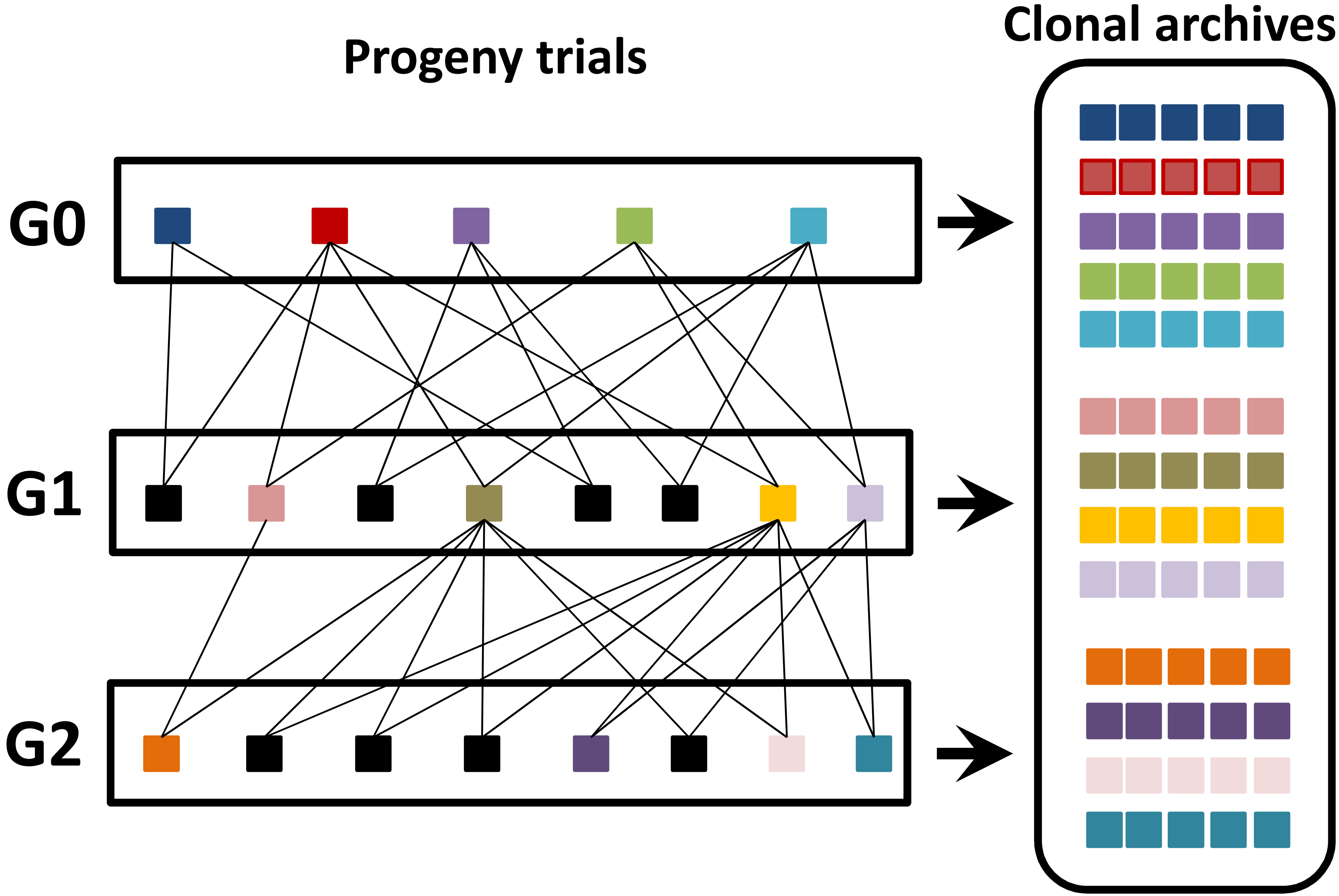
Various sources of id errors



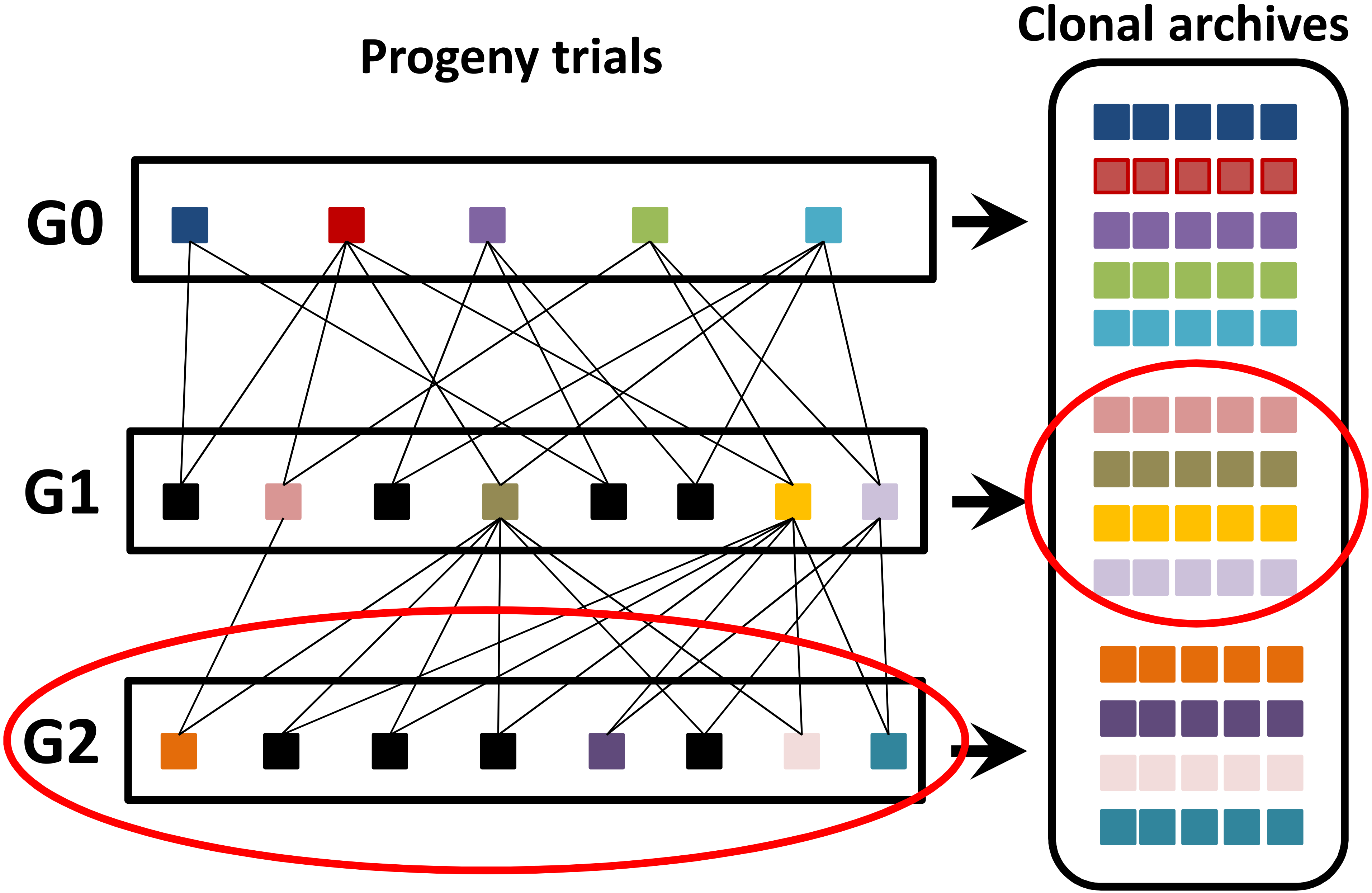
(from Marjorie Vidal PhD, 2016)

- Cumulative process over generations
- Markers = the only way to check id

Id errors in progeny trials and clonal archives



Id errors in progeny trials and clonal archives

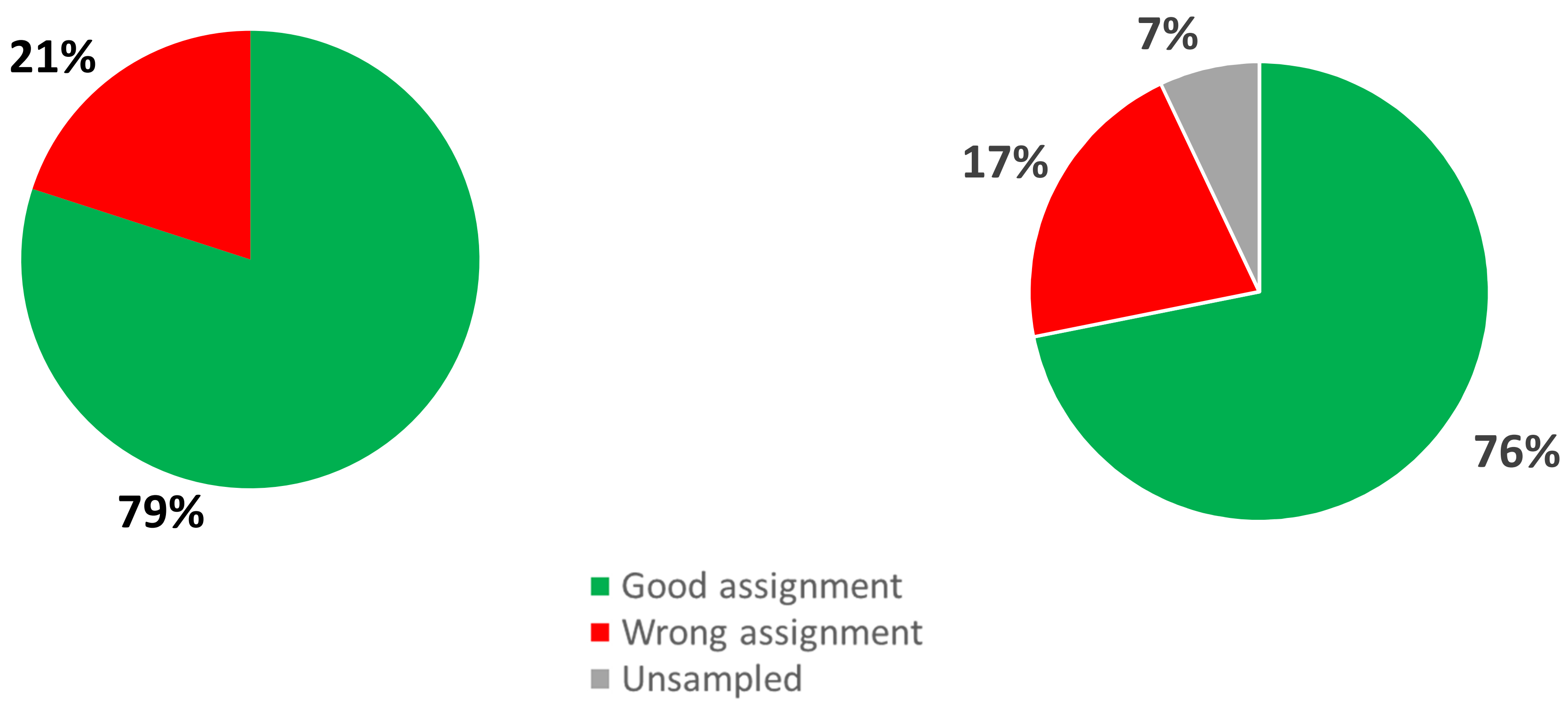


Id errors in progeny trials and clonal archives

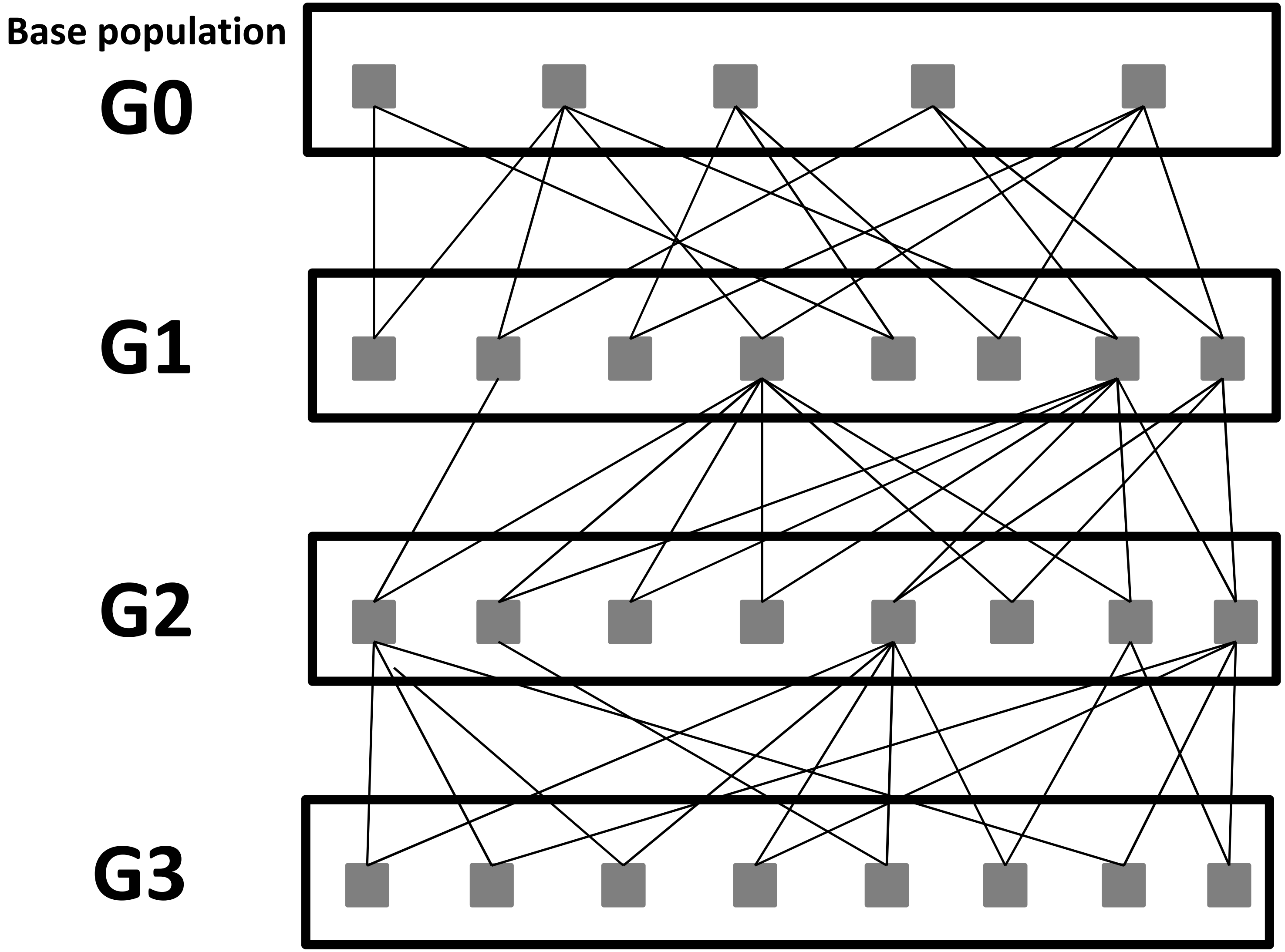
Estimation of pedigree errors in the maritime pine breeding population (80 SNPs)

G2 progeny trials
(3,009 trees)

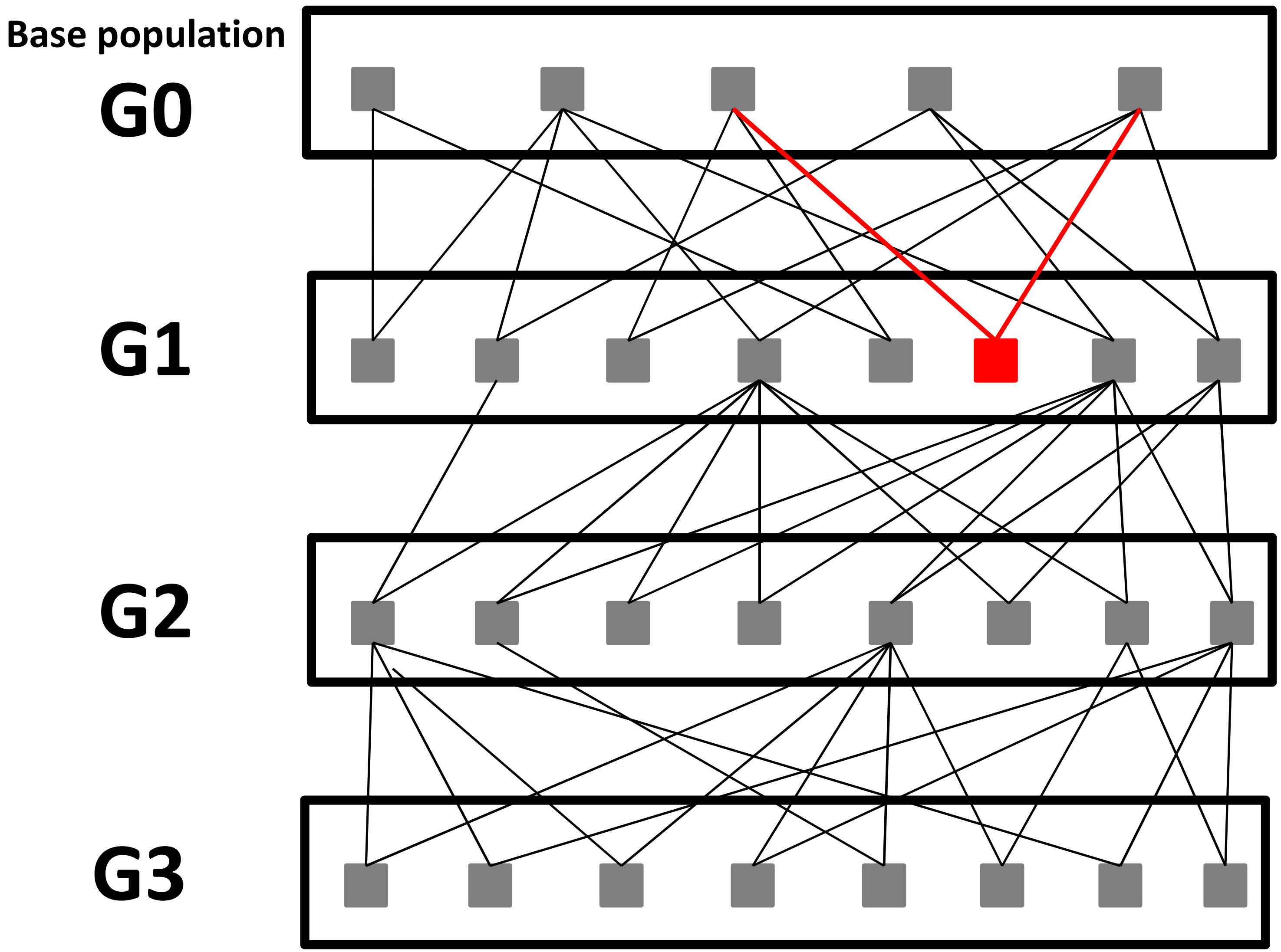
G1 clonal archives
(524 G1 trees x 5 ramets / tree)



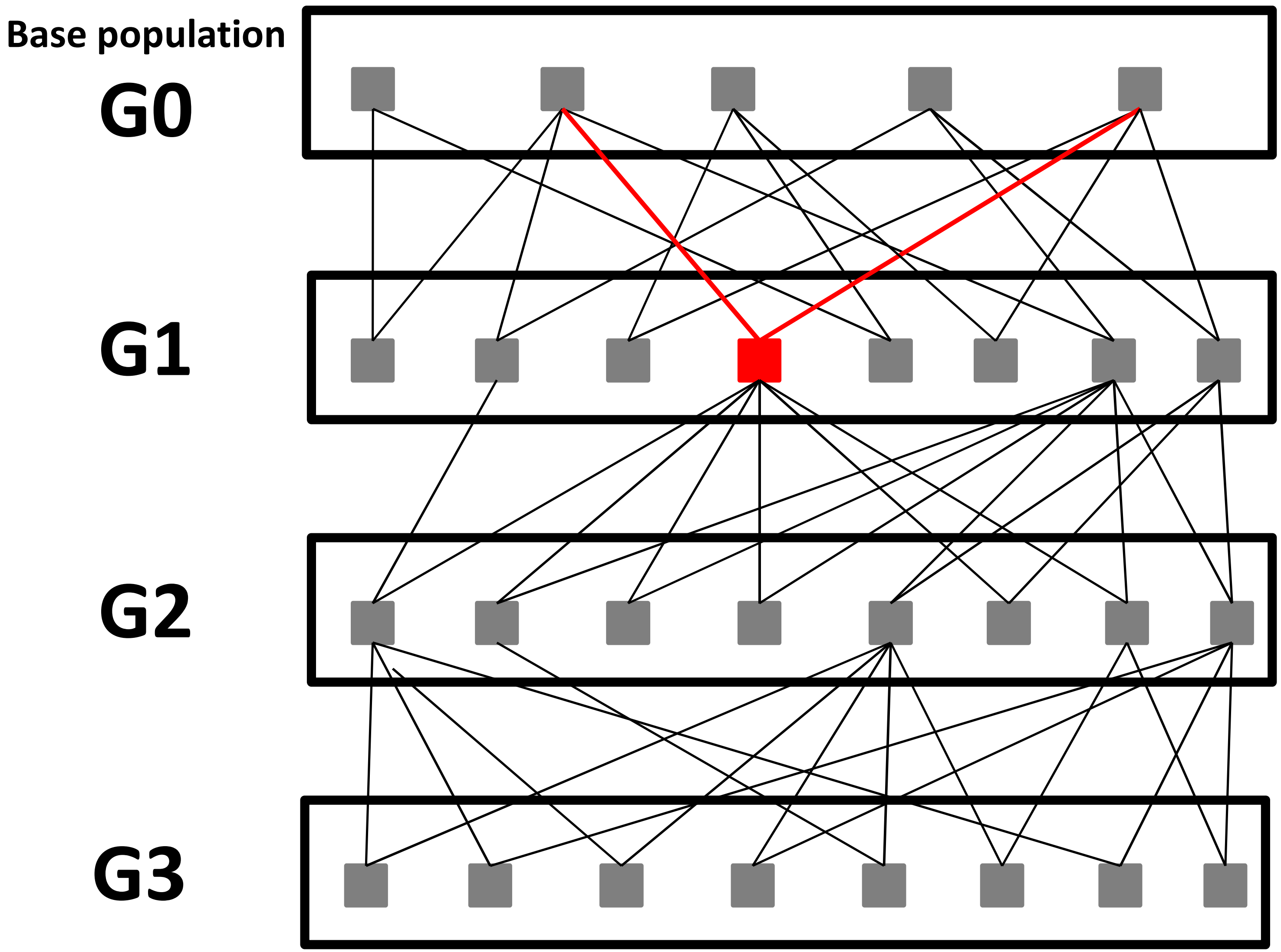
Consequences of id errors



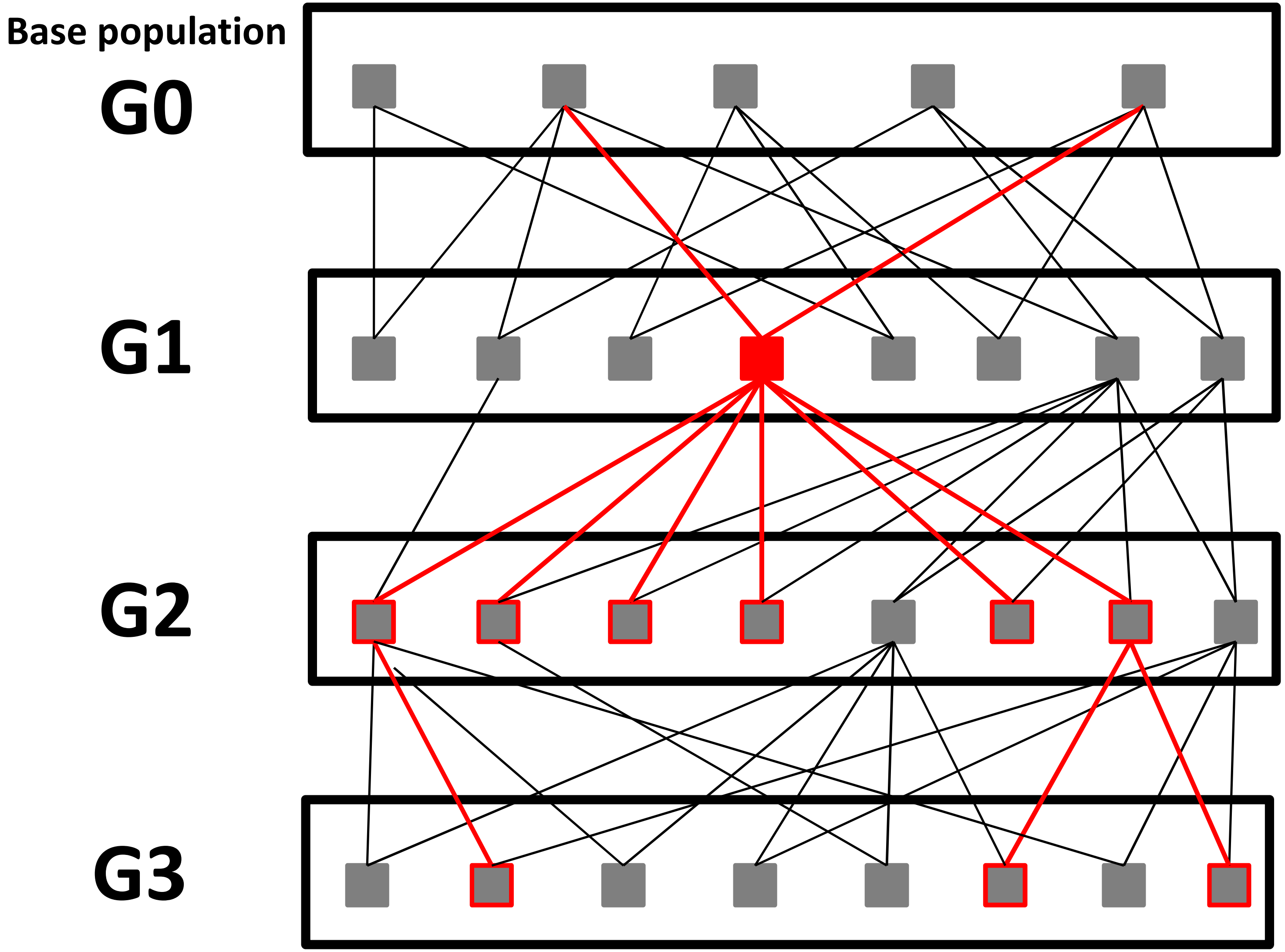
Consequences of id errors



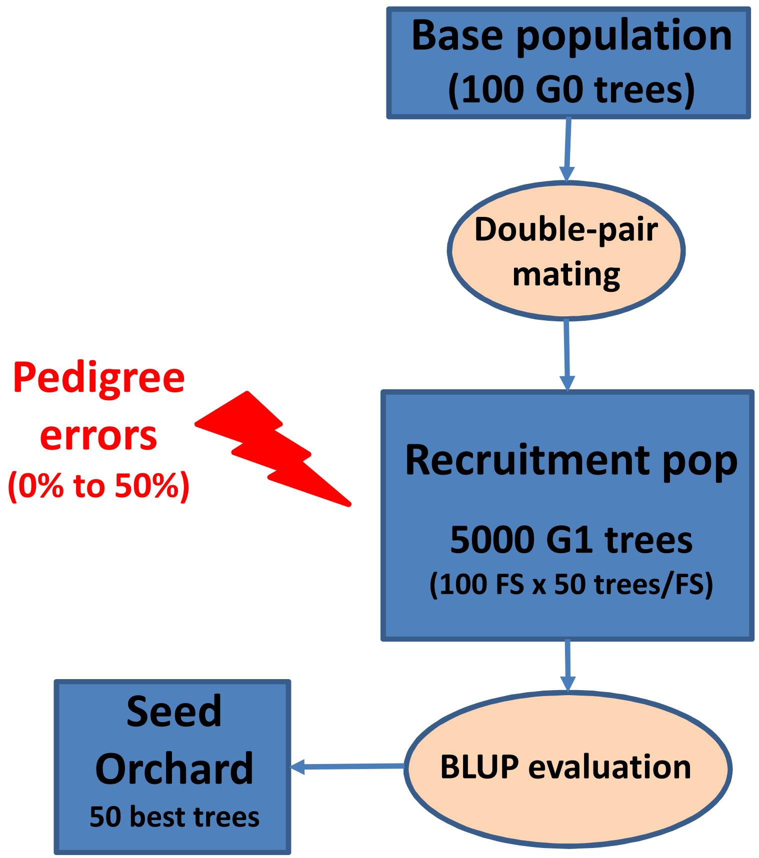
Consequences of id errors



Consequences of id errors



Evaluation of id errors consequences through simulations

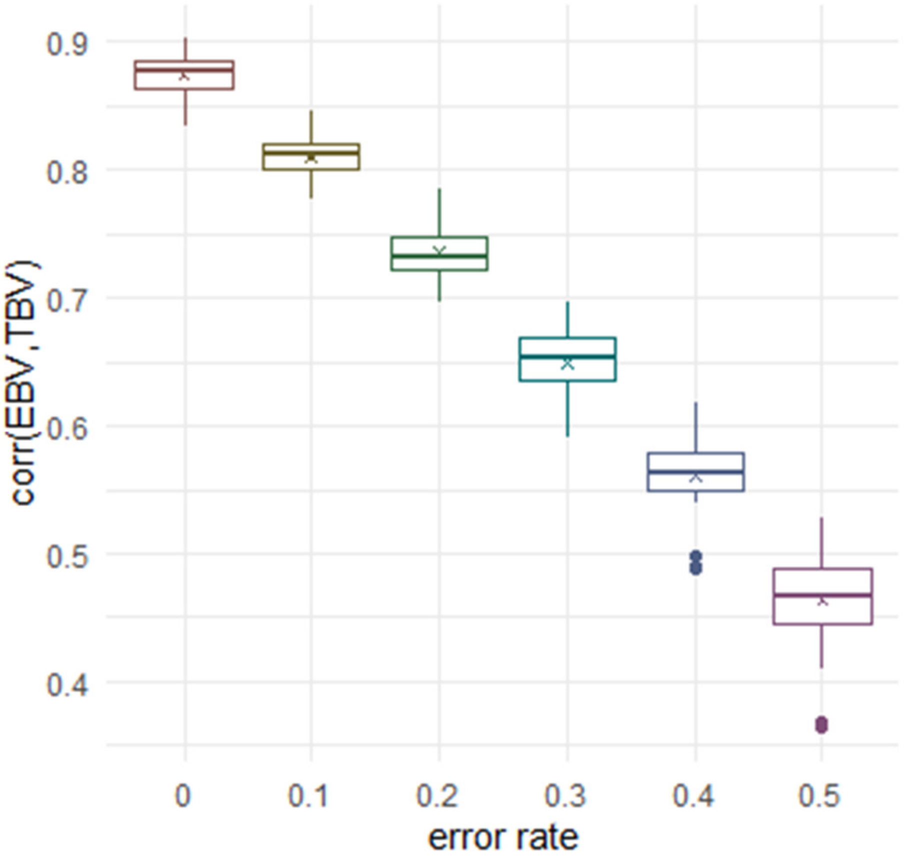


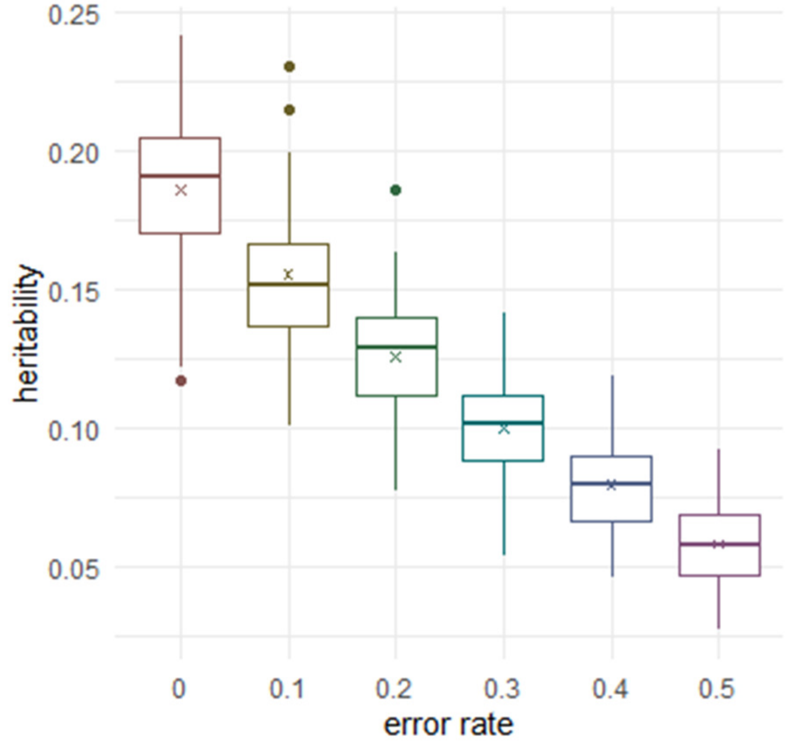
- **Stochastic simulations**
  (h^2=0.2 and CVa=15%)

- EBV from BLUP evaluation with various levels of id errors

- **Consequences on heritability, cor (TBV,EBV), genetic gain in SO?**

Evaluation of id errors consequences through simulations





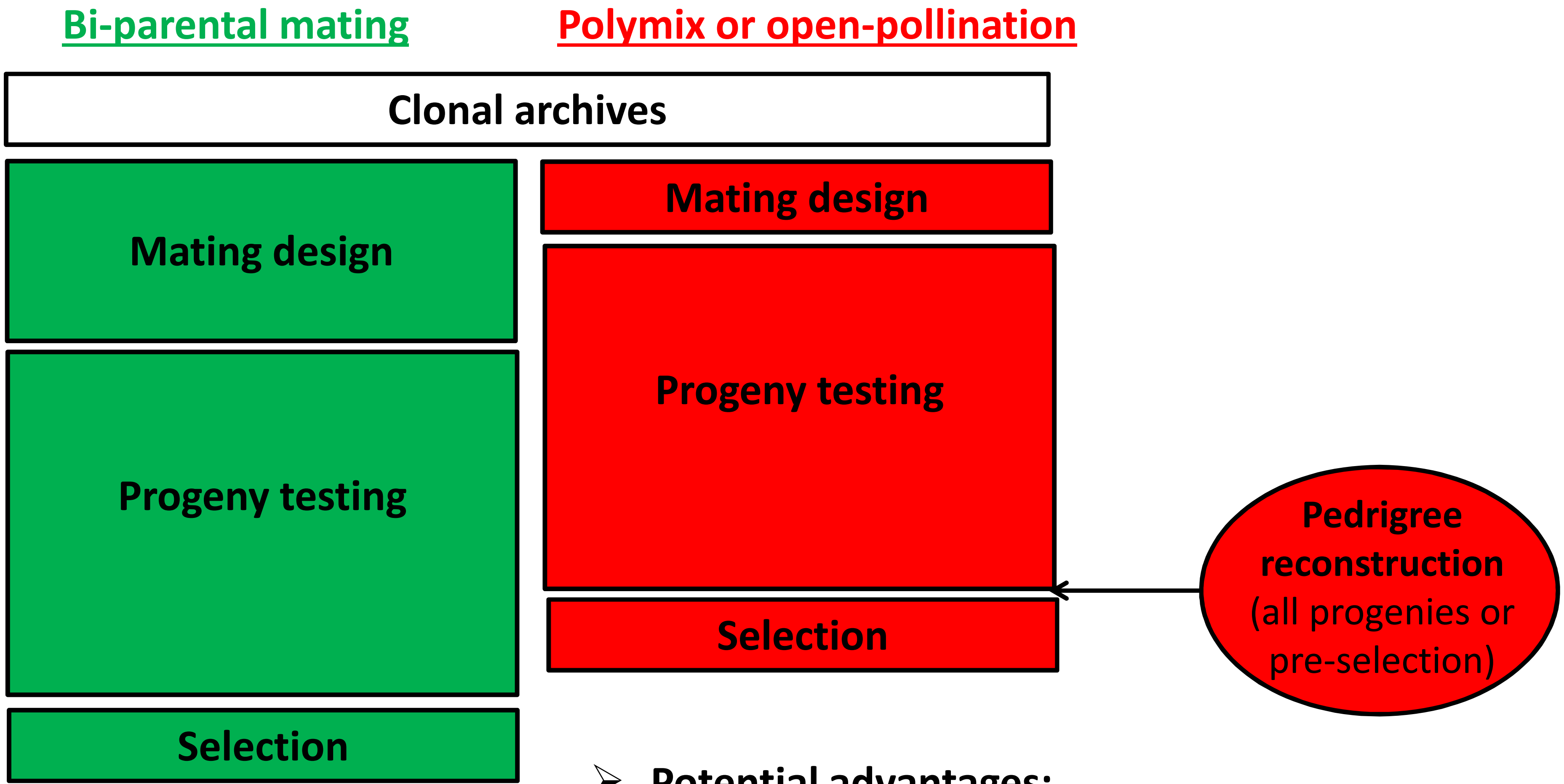
- **Higher genetic gains when the pedigree is cleansed**

« Id card » in clonal archives

- **3-year-project (2018-2020) to genotype 8,000 trees with 80 SNPs :**
  - « id card » for each genotype in clonal archives
  - correct (when possible) for pedigree mistakes in the database and in the field
  - field labelling with bar codes
- **Easier to implement during the first generations**
- **First mandatory step to go further with markers implementation in the breeding programme**

4/ Alternative mating design based on pedigree recovery

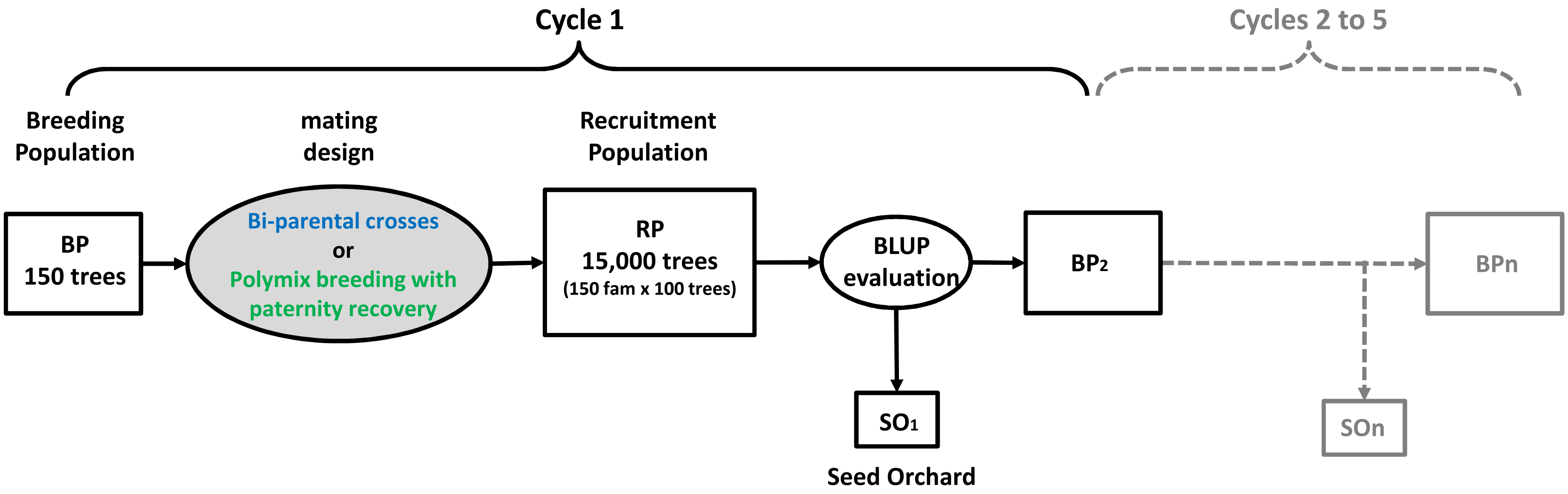
Mating design based on pedigree recovery



- **Potential advantages:**
  - Simplification of the crossing process
  - Generation of a large number of families
  - Verification of id for selected genotypes

Evaluation of mating design based on pedigree recovery through simulations

Bi-parental crosses (OC or DPM) vs. polymix breeding with pedigree recovery (PCM) (POPSIM simulator)



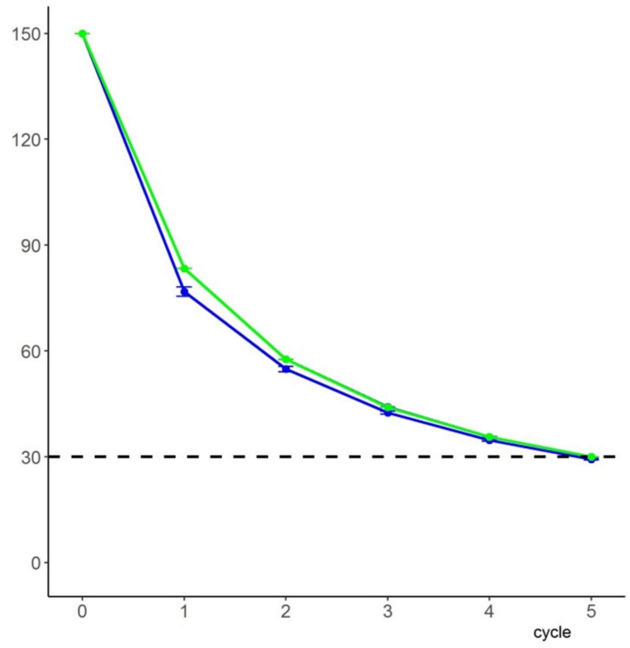
Diversity constraints:

Ns>30 afther 5 cyles in BP

Ns>10 in SO

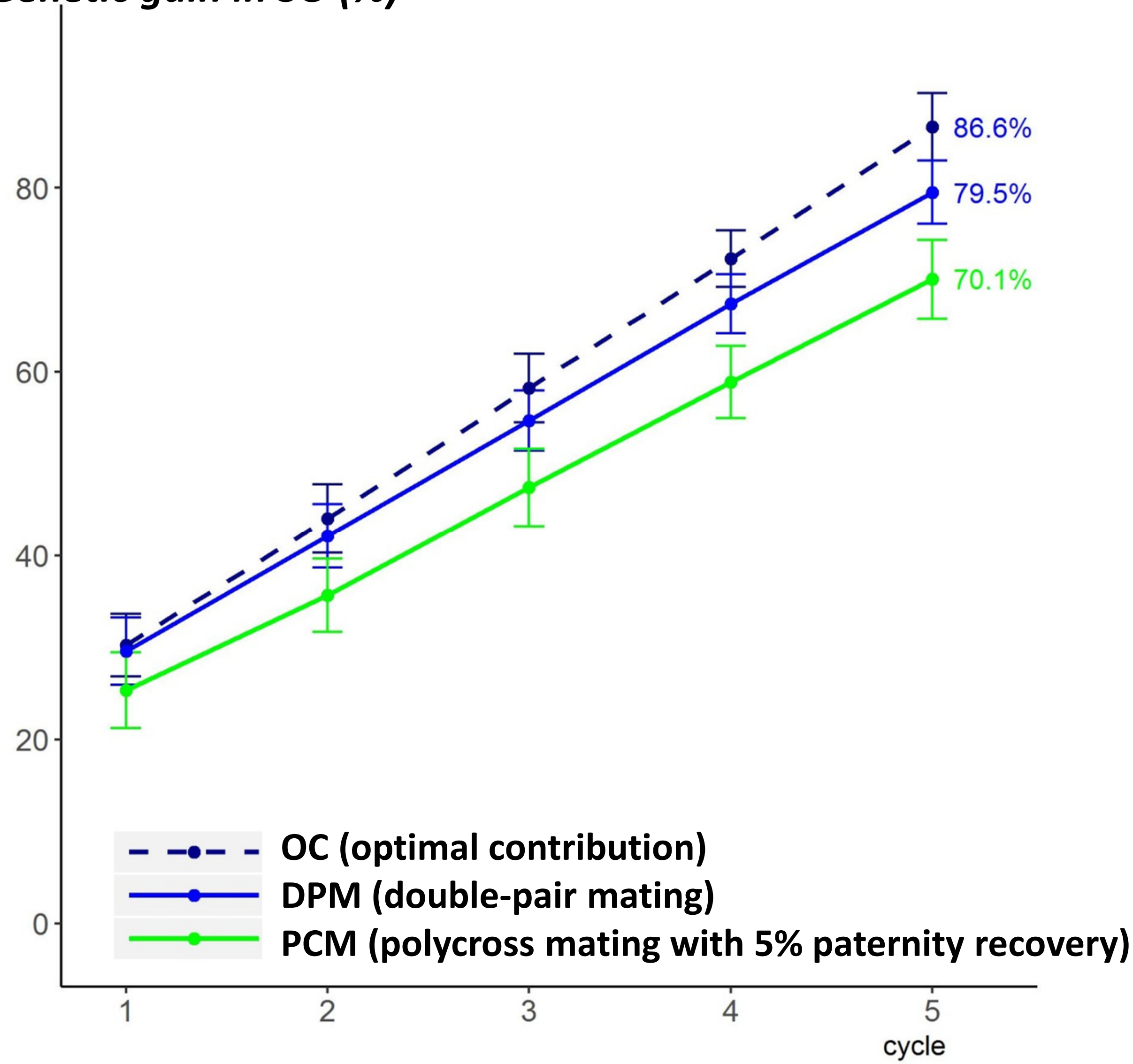
Evaluation of mating design based on pedigree recovery through simulations

Effective size (Ns) in BP



- Breeding strategies compared for a given diversity level

Genetic gain in SO (%)



Evaluation of mating design based on pedigree recovery through simulations

- DPM/OP (biparental mating) outperforms PCM (mating design based on pedigree recovery)

- Mainly due to a best management of parental contribution with bi-parental mating design (diversity constraints can be fulfilled motre easily)

- Paternity testing rate has no significant impact (it has to be sufficient to fulfill diversity constraints: 5% ie 750 trees here)

- No advantages to generate a lot of families in PMX strategy

- Superiority of DPM/OP strategy has to be mitigated based on the time required to complete the design

- Genotyping cost = extra-investment for PCM (but allows id checking at each generation)

Figure 7: Genetic gain (at breeding cycle 5) for increasing numbers of crosses

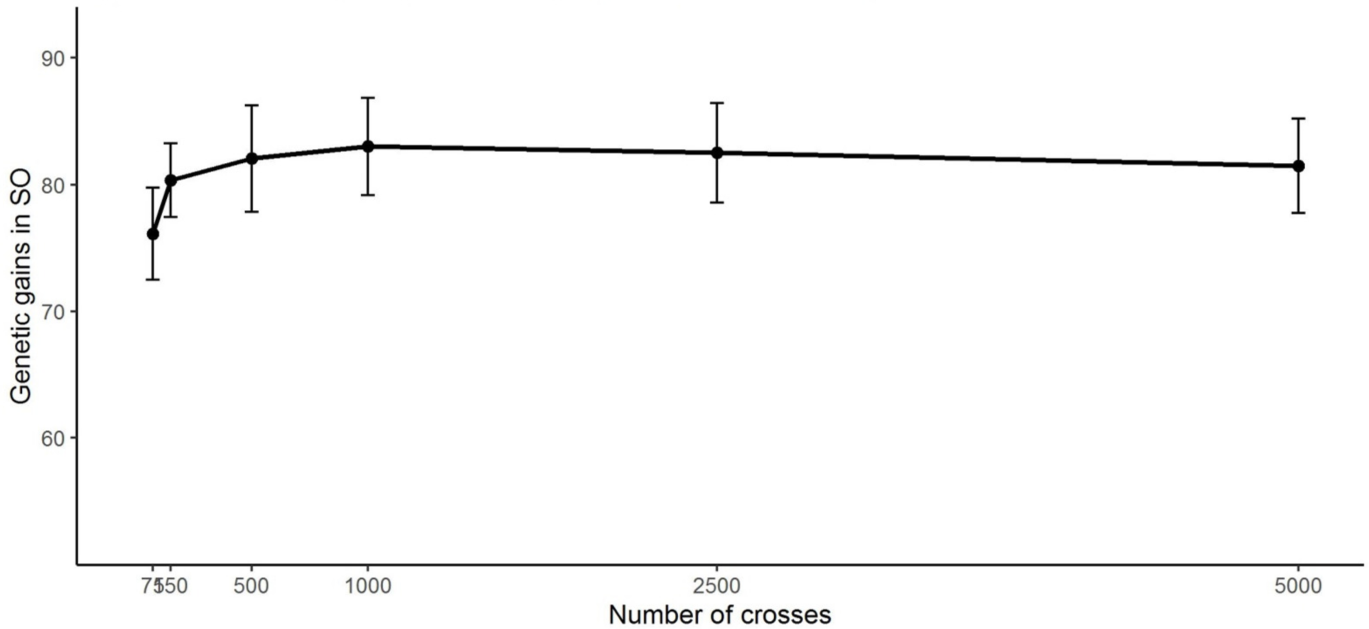
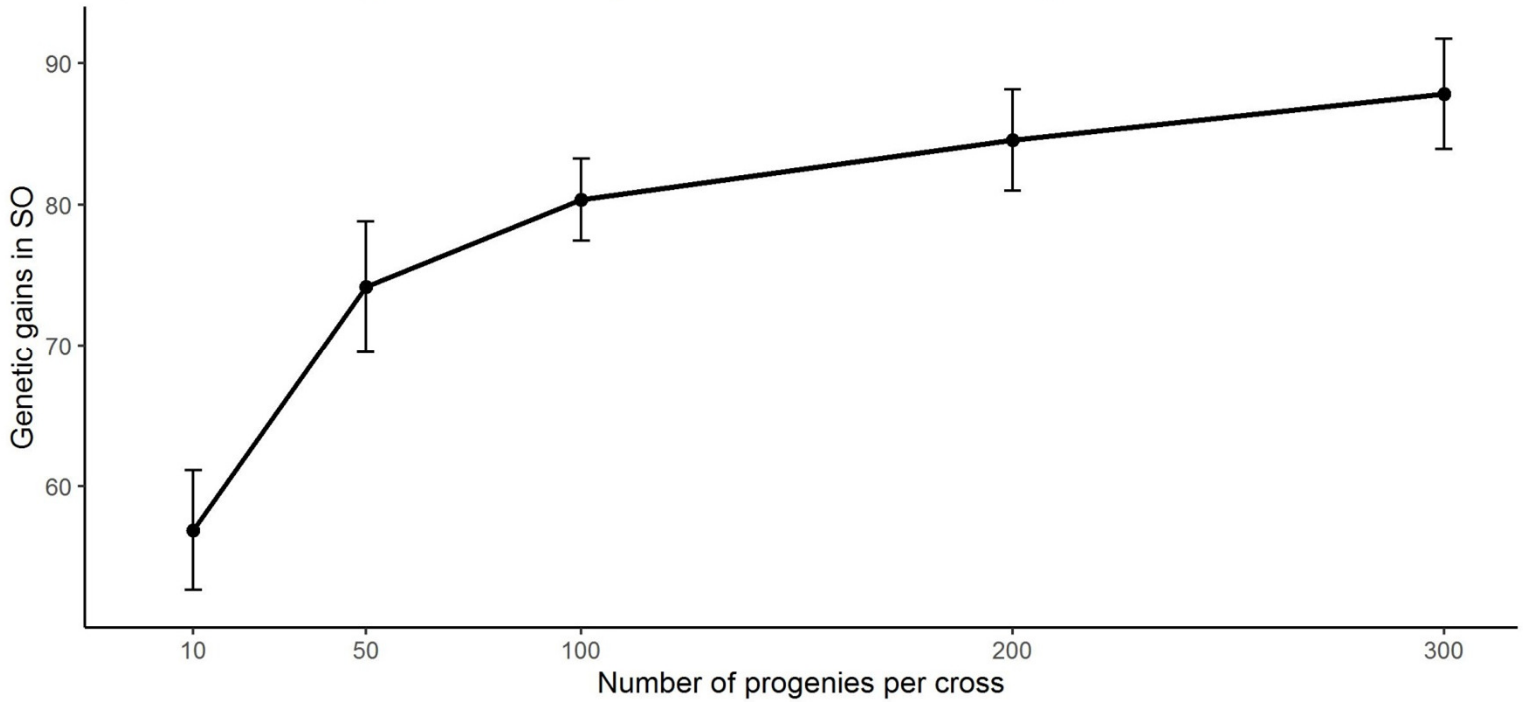


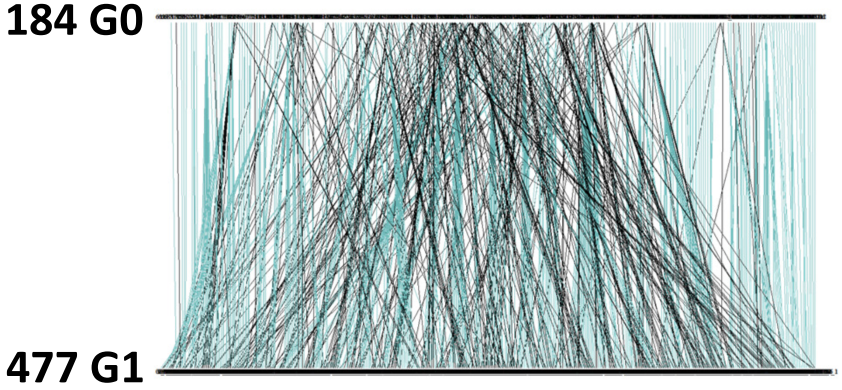
Figure 6: Genetic gain at breeding cycle 5 for increasing family sizes



5/ GS proof-of-concept in maritime pine

2 GS populations (sampled from the breeding pop)

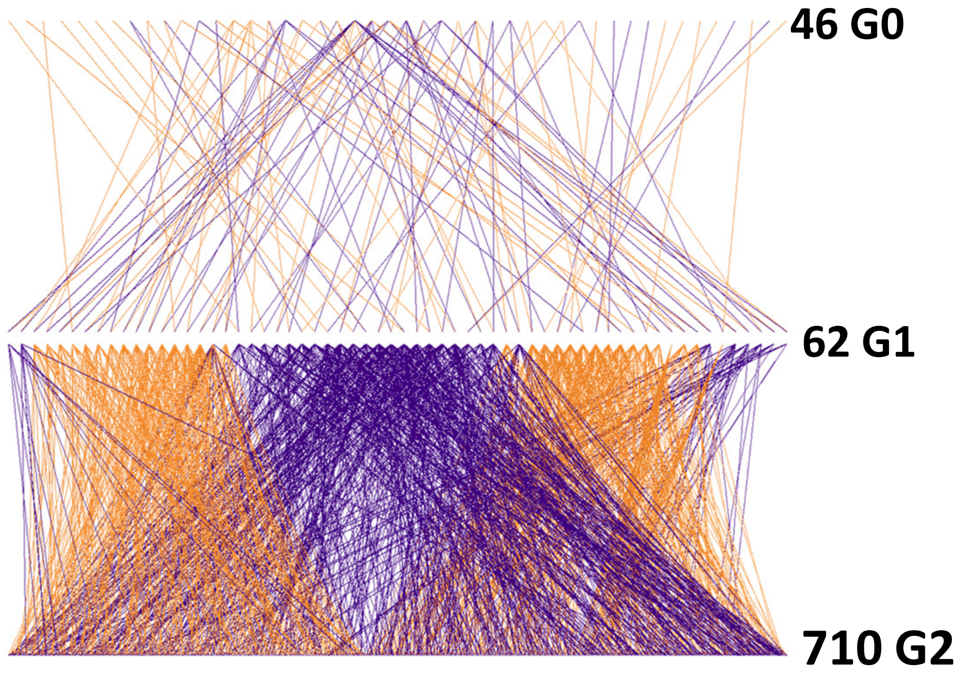
Population 1



- 661 individuals (Ne≈100)
- G0 + G1
- G1: 191 HS families (≈ 2.5 individuals)
- 2,500 SNPs
- Pseudo-phenotypes (BLUP)
- Growth, sweep

Isik *et al.* 2015

Population 2



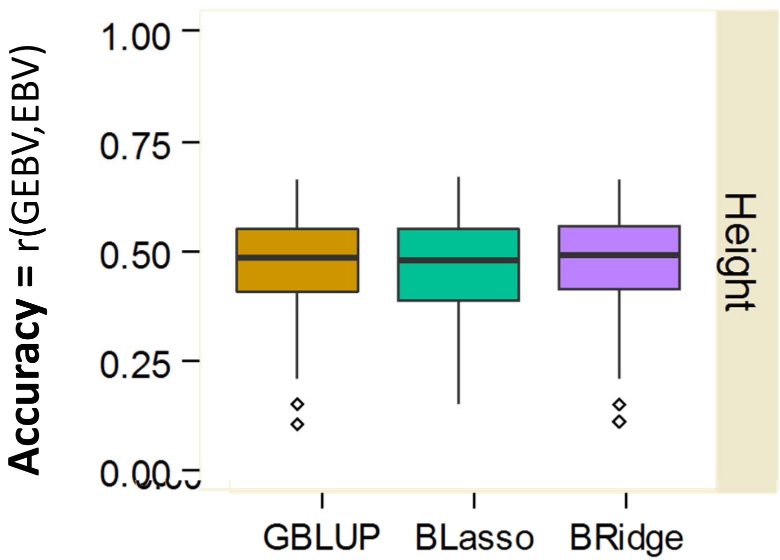
- 818 individuals (Ne=24)
- G2 + all their ancestors (G0 + G1)
- G2: 32 HS families (≈ 22 individuals)
- 4,300 SNPs
- Pseudo-phenotypes (BLUP)
- Growth, sweep

Bartholome *et al.*, 2016

Prediction accuracy

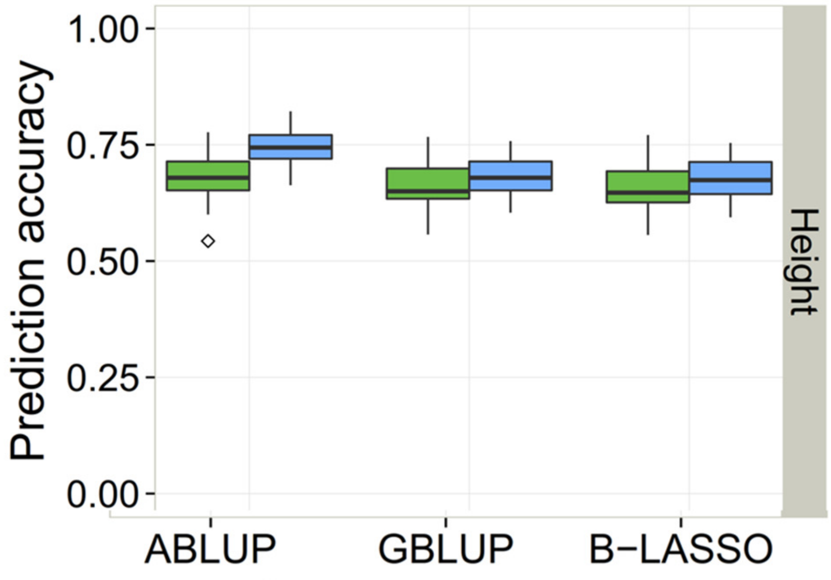
Population 1

Training: G0 + 90% G1
Validation: 10% G1



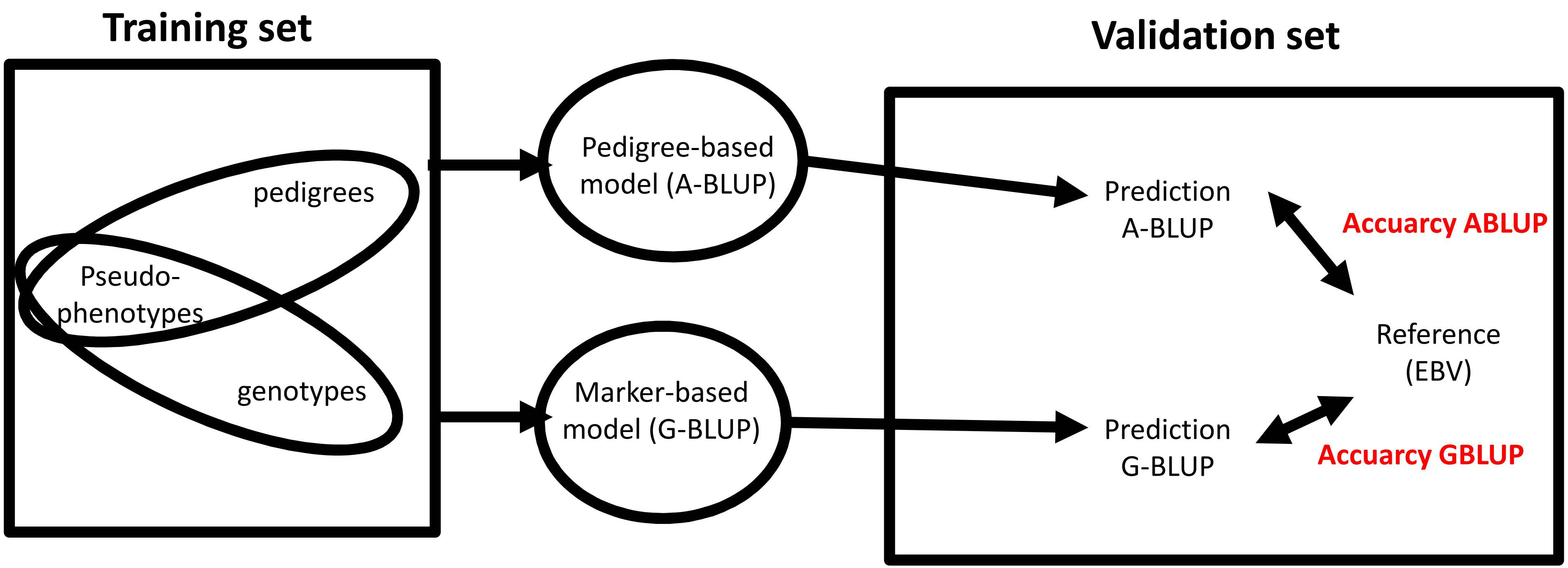
Population 2

Training: 80% G2
G0 + G1 + 80% G2
Validation: 20% G2



- Accuracy from 0.5 to 0.75 despite low linkage disequilibrium and low marker coverage of the genome (2.5 SNPs/cM in Pop2)
- Higher accuracy in Pop2 (lower effective size, all ancestors genotyped)
- Similar results whatever the method (GBLUP vs Bayesian methods)
- Similar accuracy for the pedigree-based method (A-BLUP) and marker-based methods

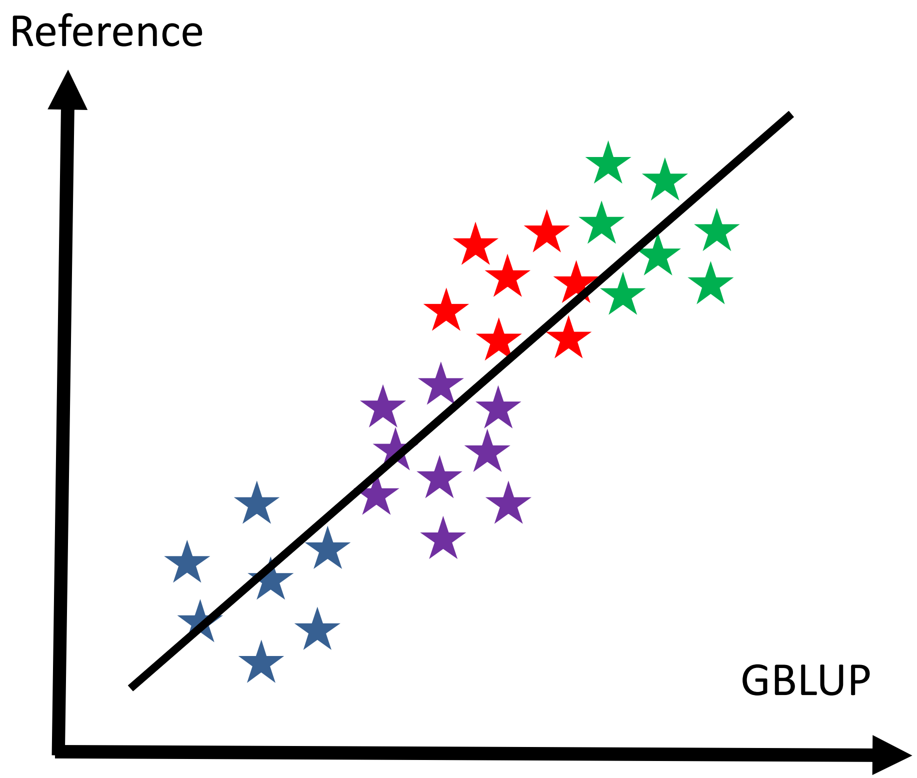
Mendelian sampling prediction: A-BLUP vs G-BLUP

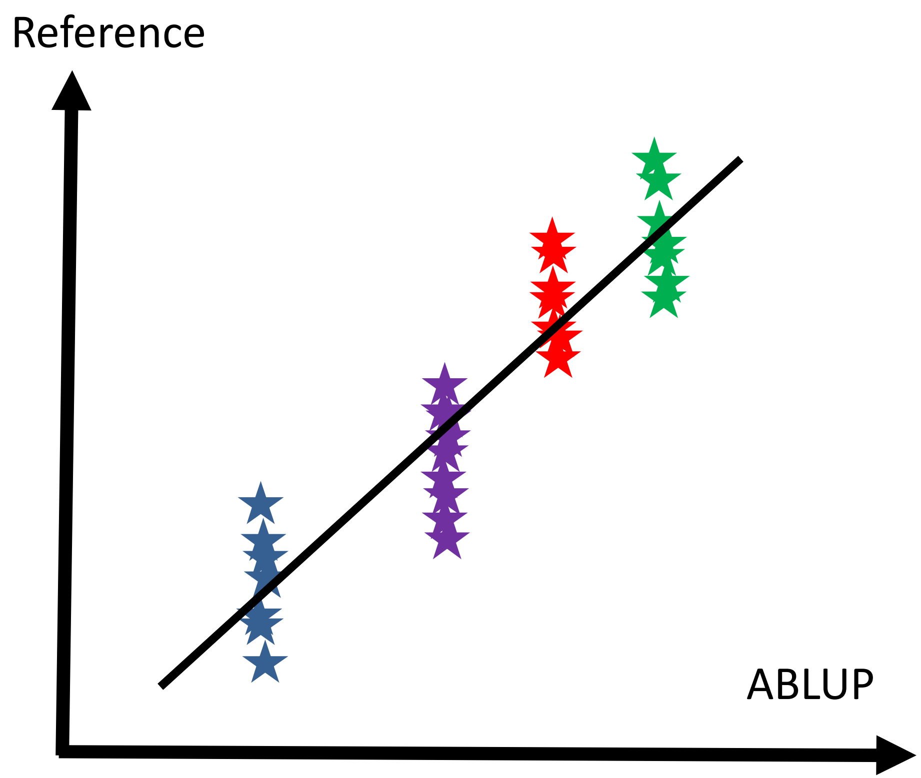


- A-BLUP evaluation: all trees from one family have the same BV if no individual phenotype / progeny (= mid-parent value)
- GS models: can predict a different BV for each tree of a family even with no phenotype (variation around mid-parent value = Mendelian sampling)

Mendelian sampling prediction: A-BLUP vs G-BLUP

GS predictive accuracy is not directly linked with the ability to predict the Mendelian sampling (GS accuracy comes from relatedness and LD markers-QTLs)





Depending on the population studied, GS models can have high accuracy even without any ability to predict Mendelian sampling

Mendelian sampling prediction: A-BLUP vs G-BLUP

- **In the litterature:**
  - Ability to predict Mendelian sampling is generally poorly discussed
  - GS accuracy mainly based on relatedness (and not LD markers-QTLs)
  - G-BLUP > A-BLUP when G-matrix allows correcting (pedigree errors) or completing kinship (full pedigree unknown, structuration in the base population not taken into account with A-BLUP)

- **For maritime pine, the similar accuracy from A-BLUP and G-BLUP can be explained by:**
  - a limited number of markers in comparison to the large size of conifer genome (more than 20 Gb)
  - a biased sampling of SNPs set (mainly SNPs from EST)
  - the reliability of the pseudo-phenotypes considered as the reference to estimate the genomic selection accuracy
  - **the design of the training population, generally with a low number of trees per family**

6/ Perspectives for GS in maritime pine



GS perspectives

A/ From pedigree based matrix (A-matrix) to markers based matrix (G-matrix)

- Keep the same BLUP methodology but substitute G-matrix (or H-matrix) for A-matrix

- Increase of accuracy as:
  - realized relationship is substituted to expected relationship (or selection earlier for the same accuracy)
  - Particularly efficient for hidden relationship (or incomplete pedigree, or structure base population)

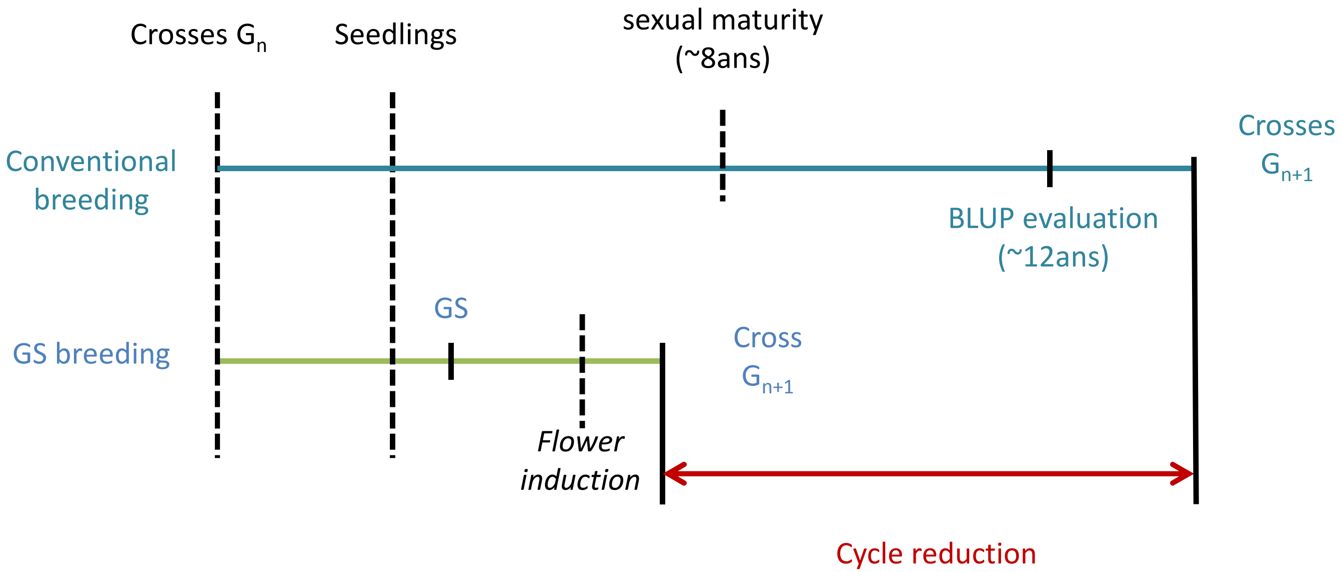
- This arises 2 issues:
  - Merge genotyping data with various nber SNP (➔ implementation)
  - Strategies to genotype the « key » trees

GS perspectives

B/ Selection without phenotyping (decrease breeding cycle length)

- The more attractive perspective but depends on the sexual maturity age (or associated with flowering induction)

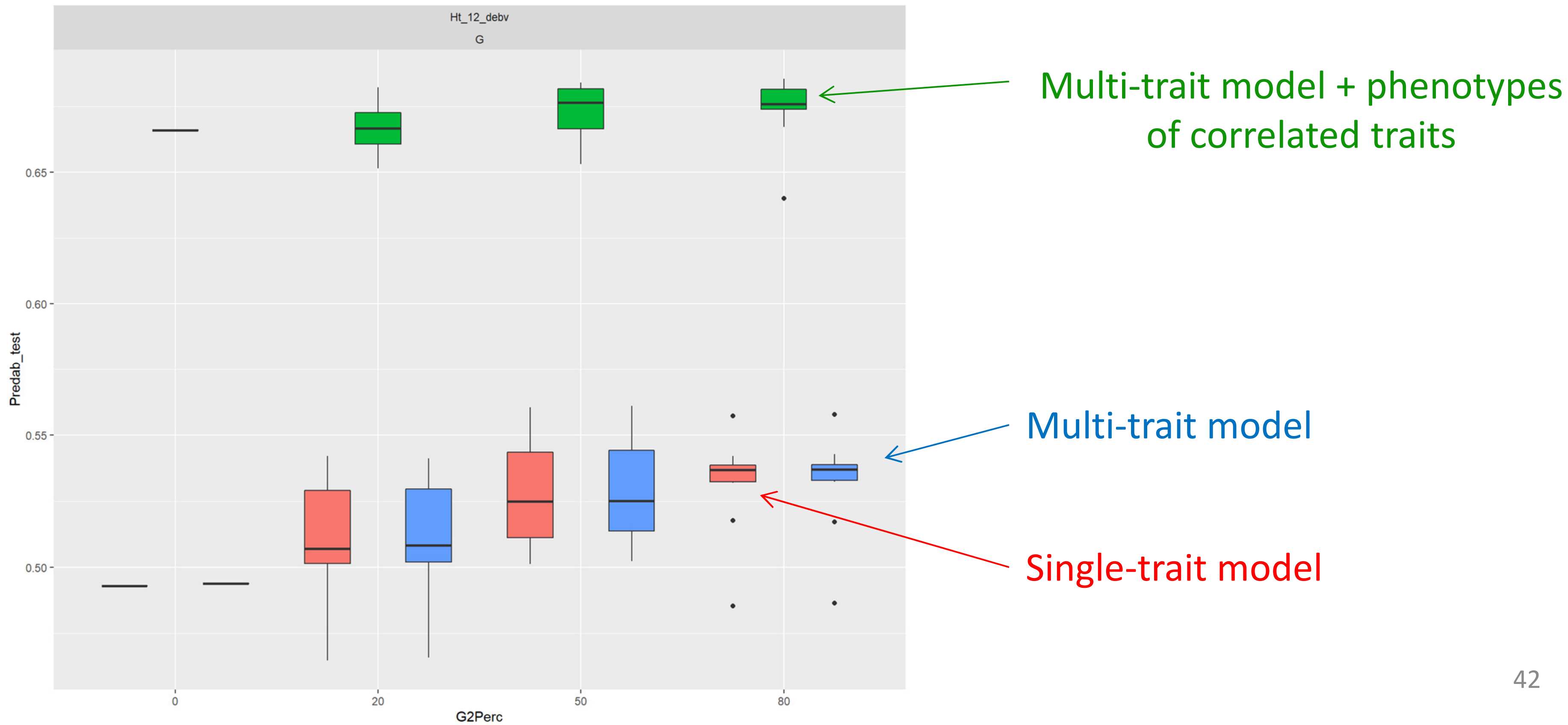
- Can be used for pre-screening (ie selection intensity increase)



GS perspectives

C/ Selection for costly traits

- For instance: pest resistance, drought resistance…
- Phenotyping focused in the training population
- Multitrait evaluation



Conclusion

- **Main GS objectives in maritime pine:**
  1/ selection for costly traits (disease resistance, climate change)
  2/ shortening of breeding cycle duration

- **Main research area in GS:**
  ability to predict the variability within families i.e. Mendelian sampling (design of the training population, phenotypes reliability, number of markers)

- **Optimal breeding strategy vs. the one we can effectively apply!**
  - GS implies: new field work, new databases, new skills…
  - implementation of markers as a first step (id checking, pedigree recovery)

- **Even without GS, use of markers increases genetic gains and convinces scepital people about the use of markers in breeding!**

Thanks for your attention!

Acknowledgment

biogeco

Christophe Plomion, Jérôme Bartholomé, Marjorie Vidal, Marc Kleinhentz, Marie Pégard, Geoffrey Haristoy

cgfb
GÉNOME TRANSCRIPTOME

Erwan Guichoux, Christophe Boury, Marie Massot

UNITÉ EXPÉRIMENTALE FORÊT PIERROTON

Annie Raffin, Patrick Pastuszka, Frédéric Bernier

Groupe Pin Maritime du Futur