Use of molecular markers to enhance genetic gains in the maritime pine breeding program

Laurent Bouffier, Marjorie Vidal, Jérôme Bartholomé, Christophe C. Boury,
Annie A. Raffin, Christophe Plomion

Use of molecular markers to enhance genetic gains in the maritime pine breeding program

Laurent Bouffier, Marjorie Vidal, Jérôme Bartholomé, Christophe Boury, Annie Raffin, Christophe Plomion

Landscapes in southwestern France

French vineyards

Forest area

0 - 15 %   15 - 45 %   45 - 100 %
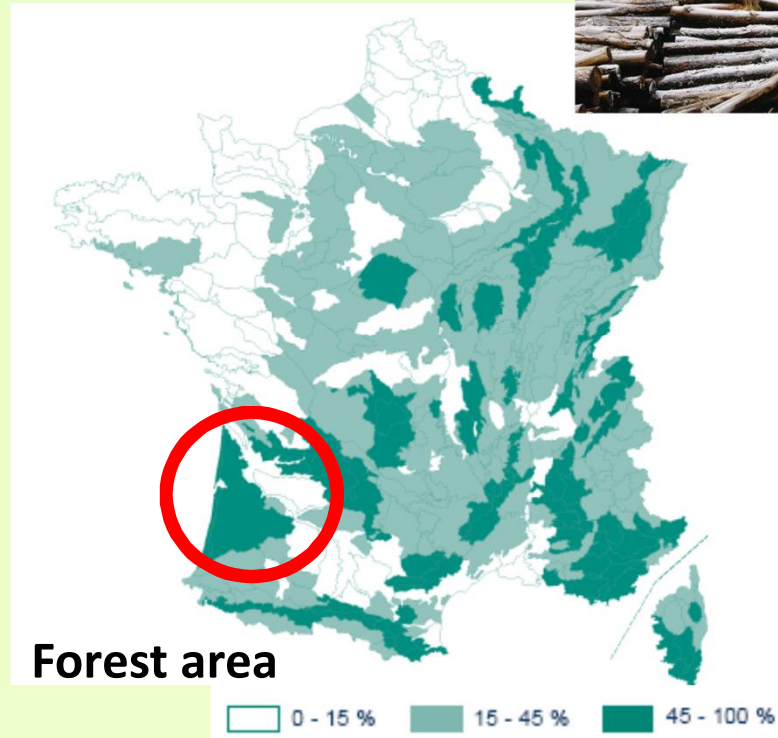
Maritime pine forest in southwestern France

Les Landes de Gascogne
- Massifs dunaires de la côte
- Forêt des Landes de Gascogne

➢ **Maritime pine forest established at the end of the 19th century**

➢ **0.8 million hectares forest**
  - Maritime pine plantations
  - 24% of French wood harvest

➢ **A breeding program since the early 1960s**
  - base population selected in the Landes forest
  - three generations: G0, G1, G2

➢ **More than 20 millions improved seedlings produced per year**
  - open-pollinated seed orchards (one breeding zone)
  - genetic gains : +30% for growth ; +30% for stem sweep

Recurrent selection scheme

Breeding population

Deployment population

Plus trees selection
= G0 population

Full-sib crosses

Polycross trials

« forward » selection

« backward » selection

Selection of best progenies = G1

Selection of best parents

Seed orchards

Full-sib crosses

Polycross trials

Selection of best progenies = G2

Selection of best parents

Seed orchards
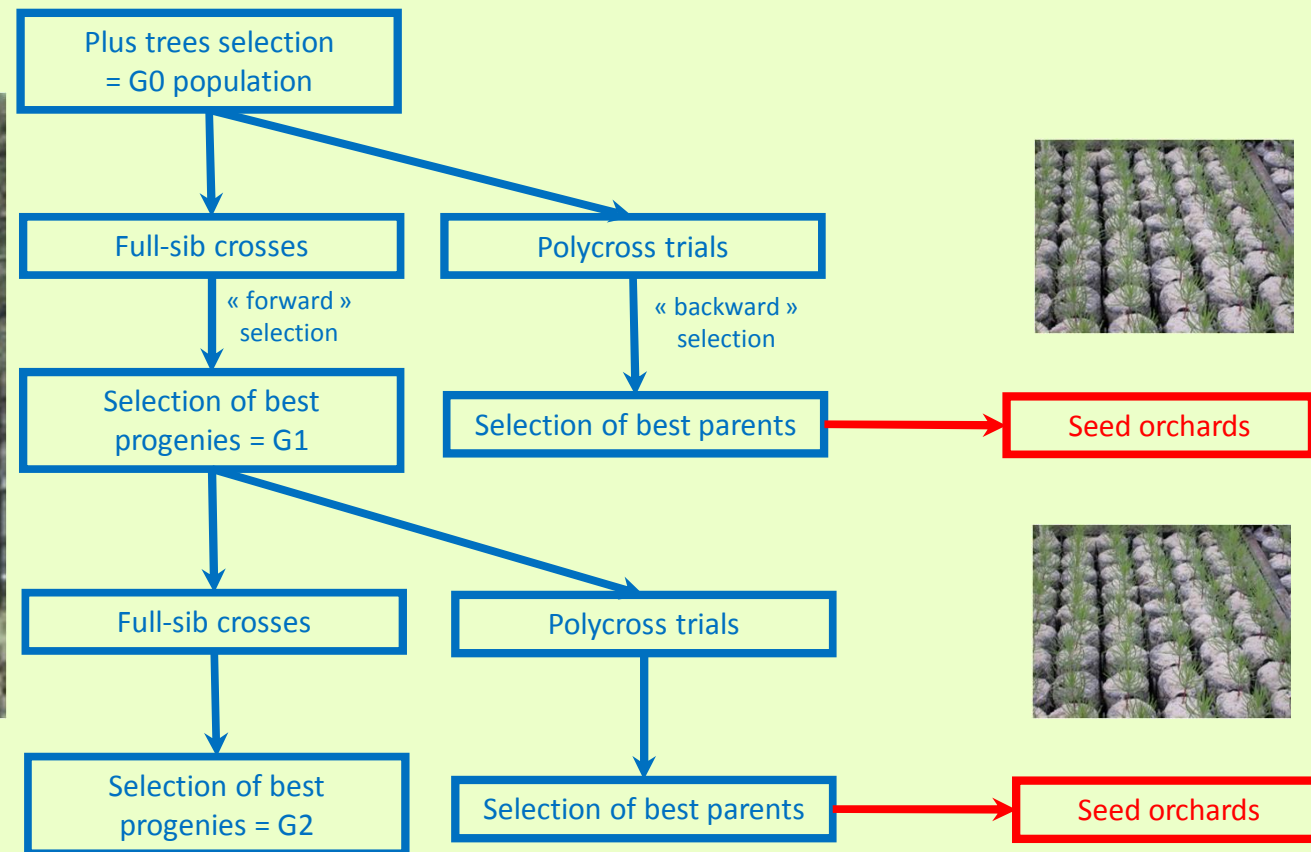
Genetic evaluation

Objective = BV comparable over trials and generations
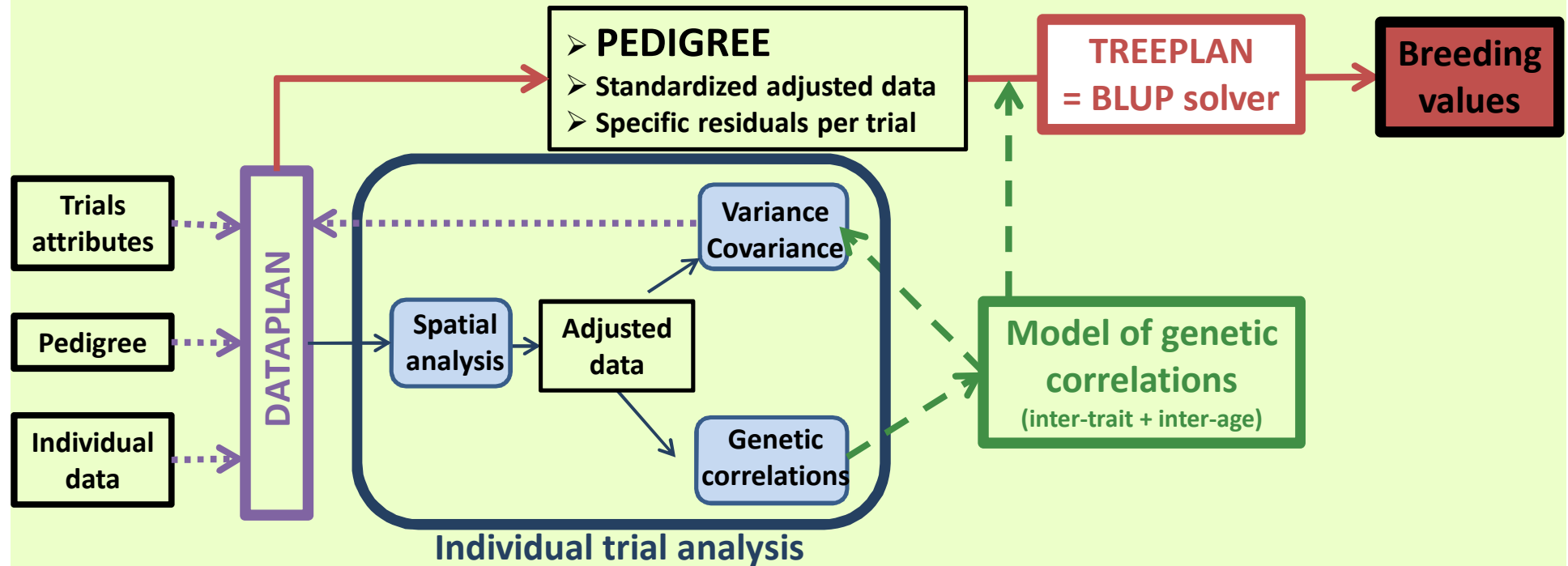
Methods =   - BLUP analysis
- taking advantage of:
  - ✓ pedigree connections (>500,000 trees)
  - ✓ correlations between traits and ages

Implementation of molecular markers

➢ A 9k SNP array available for maritime pine (Plomion *et al.,* 2016)

Operational	″ **1/ Identities / pedigree control**
	➢ DNA fingerprinting
Short term	″ **2/ Pedigree recovery to simplify crosses**
	➢ polymix breeding with pedigree recovery
Mid-term	″ **3/ Genomic selection**
	➢ breeding values prediction

1/ Identities / pedigree control

➢ dedicated SNP array to control identities (Vidal *et al.*, 2015)

5,652 polymorphic SNP

169 SNP with MAF > 0.45

121 SNP with LD<0.3

80 SNP
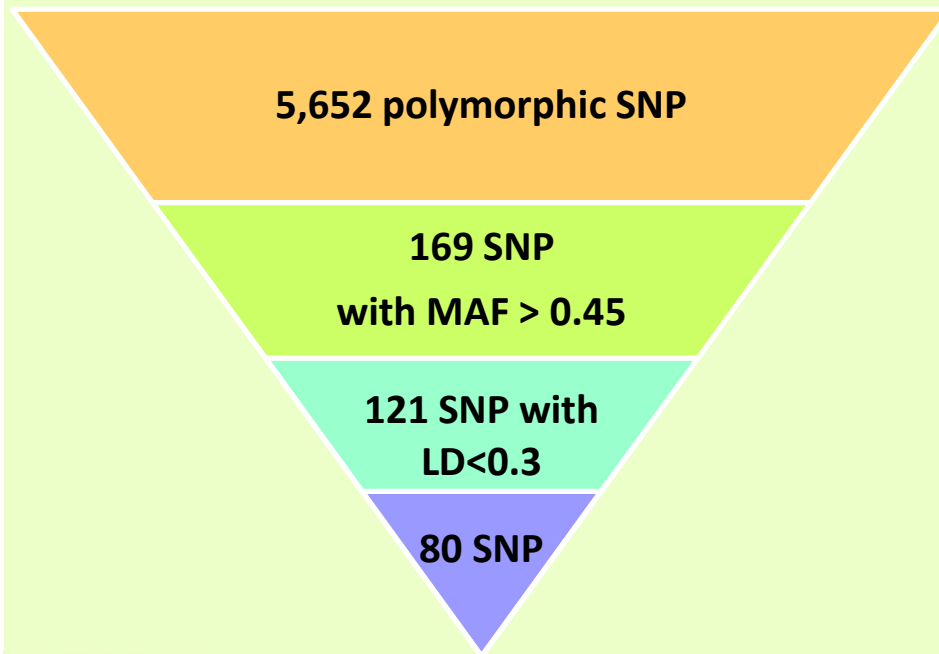
DNA genotyping (Sequenom)

« identity card »

➢ Pedigree checking with Cervus software (P> 99%)

one parent is wrong

9% (51)

18% (100) — one parent not sampled

73% (401)

both parents are correct
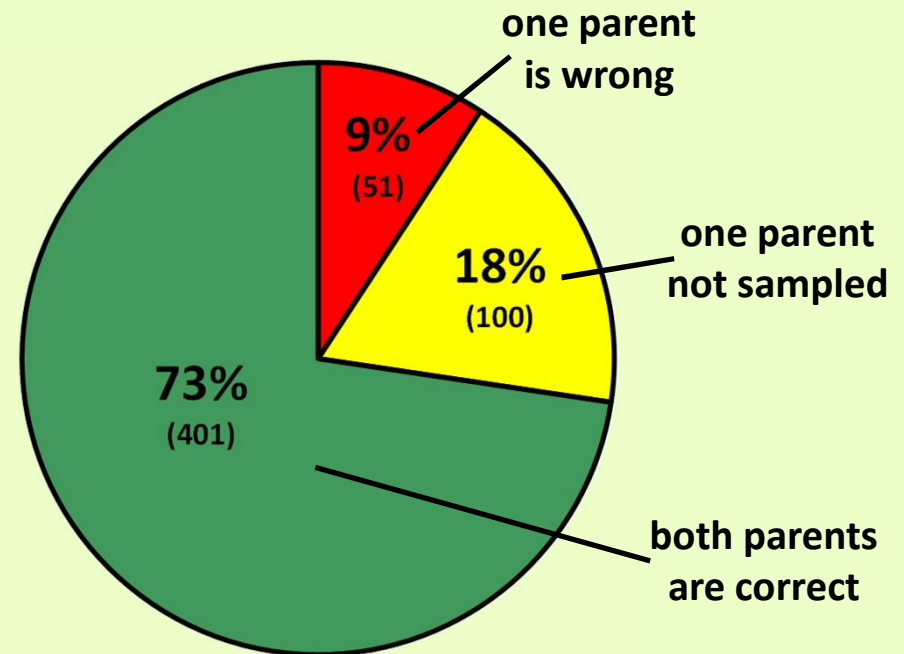
Analysis to check pedigree of 552 G1

1/ Identities / pedigree control

```
┌─────────────────────────┐
│   Collecting grafts      │  (E₁)
│   on selected trees      │
└─────────────────────────┘
           ↓
┌─────────────────────────┐
│   Grafting in nursery    │  (E₂)
└─────────────────────────┘
           ↓
┌─────────────────────────┐
│  Planting in clonal archives │  (E₃)
│ (several ramets per genotype)│
└─────────────────────────┘
           ↓
┌─────────────────────────┐
│         Crosses          │
│  Pollen  X  Female       │  (E₄)
│             flowers      │
│ ┌──────────┐             │
│ │Collecting,│            │
│ │ drying,   │ ┌─────────┐│
│ │ storing   │ │Isolation││
│ └──────────┘ └─────────┘│
└─────────────────────────┘
           ↓
┌─────────────────────────┐
│ Collecting cones, drying, seeds │  (E₅)
│   extraction, storing    │
└─────────────────────────┘
           ↓
┌─────────────────────────┐
│   Seedling in nursery    │  (E₆)
└─────────────────────────┘
           ↓
┌─────────────────────────┐
│ Establishment of progeny trial │  (E₇)
└─────────────────────────┘
```

- Collecting grafts on selected trees E_1
- Grafting in nursery E_2
- Planting in clonal archives (several ramets per genotype) E_3
- Crosses — Pollen X Female flowers; Collecting, drying, storing; Isolation E_4
- Collecting cones, drying, seeds extraction, storing E_5
- Seedling in nursery E_6
- Establishment of progeny trial E_7

➢ Various sources of identity errors

➢ It can be difficult to detect the error origin

➢ Cumulative process over generations

(from Marjorie Vidal PhD, 2016)

1/ Identities / pedigree control

➤ **Objectives for the maritime pine breeding population:**
- ✓ molecular identity for each genotype
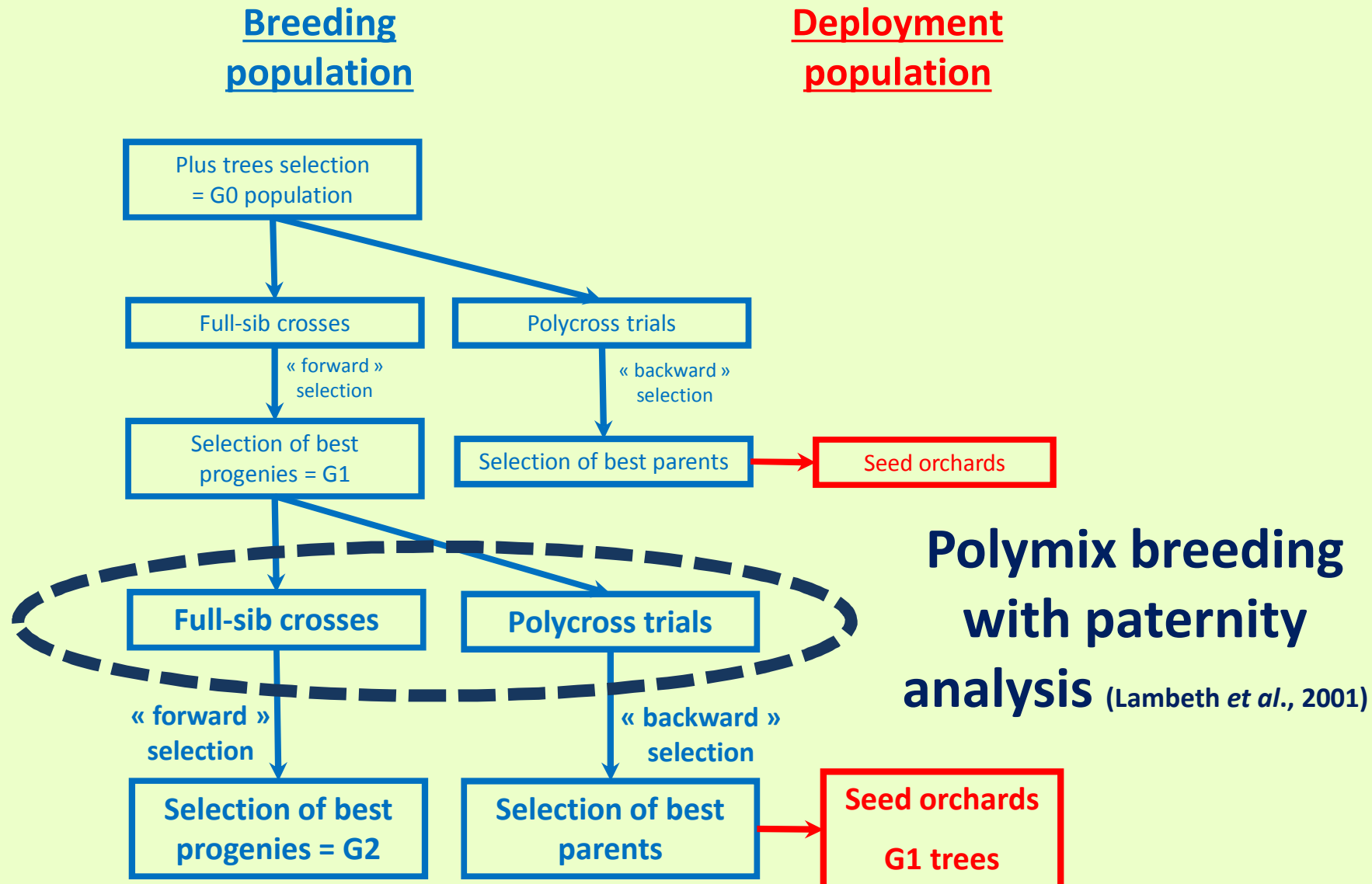- ✓ pedigree validation

➤ **Methods:**
- ✓ genotyping of clonal archives (all ramets per genotype)
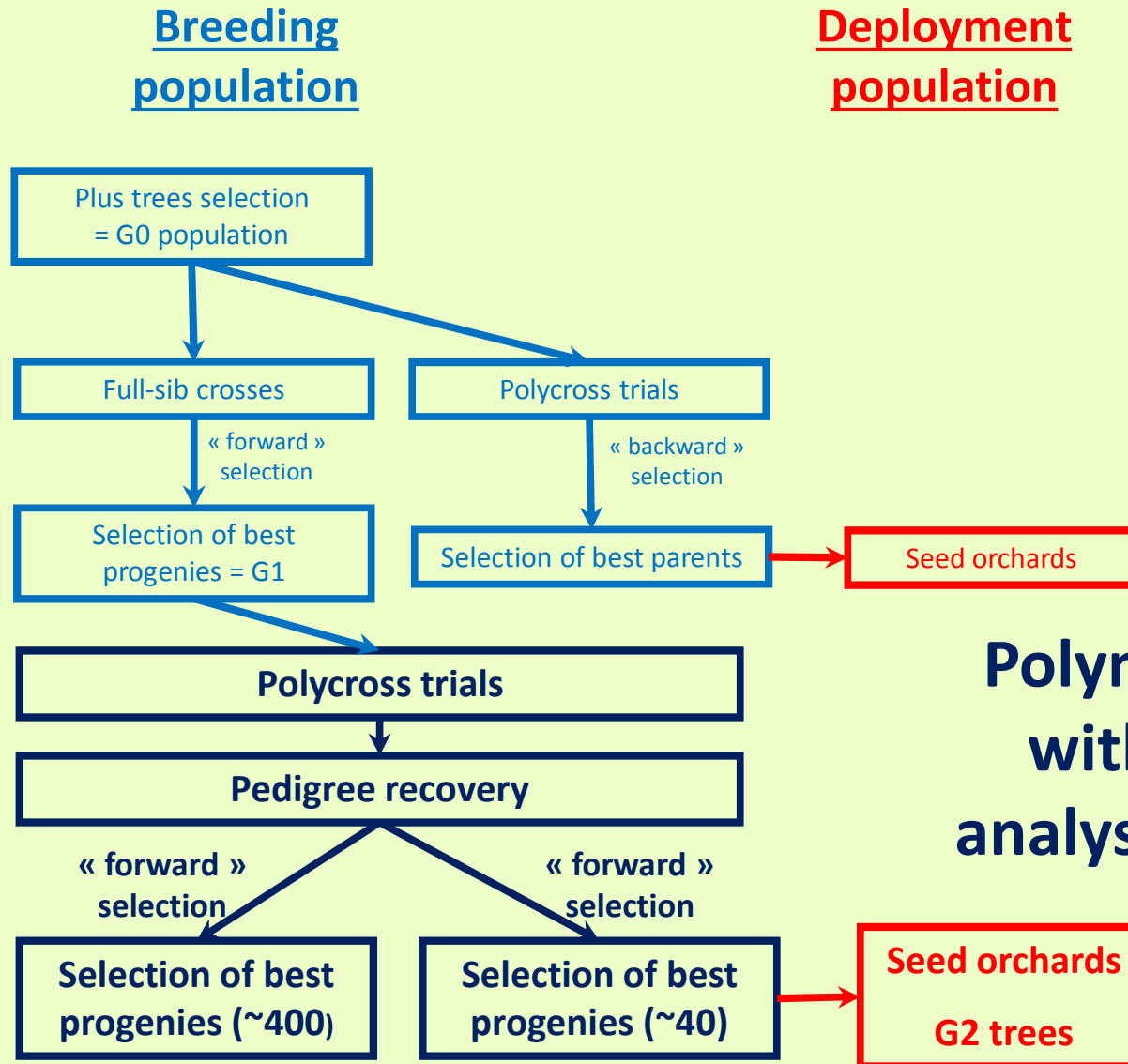- ✓ starting in 2016 (several thousands of trees to genotype)

➤ **Id / pedigree control will increase genetic gains:**
- ➔ **higher accuracy for breeding values**
- ➔ **exclude genotype errors for the deployment**

2/ Polymix breeding with paternity recovery

Breeding population

Deployment population

Plus trees selection
= G0 population

Full-sib crosses

Polycross trials

« forward » selection

« backward » selection

Selection of best progenies = G1

Selection of best parents → Seed orchards

Full-sib crosses

Polycross trials

Polymix breeding with paternity analysis (Lambeth *et al.*, 2001)

« forward » selection

« backward » selection

Selection of best progenies = G2

Selection of best parents → **Seed orchards G1 trees**

2/ Polymix breeding with paternity recovery

Breeding population

Deployment population

Plus trees selection = G0 population

Full-sib crosses

Polycross trials

« forward » selection

« backward » selection

Selection of best progenies = G1

Selection of best parents

Seed orchards

Polycross trials

Pedigree recovery

« forward » selection

« forward » selection

Selection of best progenies (~400)

Selection of best progenies (~40)

Seed orchards

G2 trees

Polymix breeding with paternity analysis (Lambeth *et al.*, 2001)

Evaluate the feasability of polymix breeding with paternity recovery

174 G1

half-sibs families

Polymix (43 G1   )

35 trees / family + checklots ➔ 6,440 trees

➢ Estimate the efficiency to maximize recombination between parents

➢ Estimate the gentic gains for deployment based on forward selection

2/ Polymix breeding with paternity recovery

➢ All fathers of the polymix contribute to the progenies (75% within the expected range)

Confidence interval

Expected mean contribution

Paternal contribution

43 fathers (polymix)

Vidal *et al.* 2015
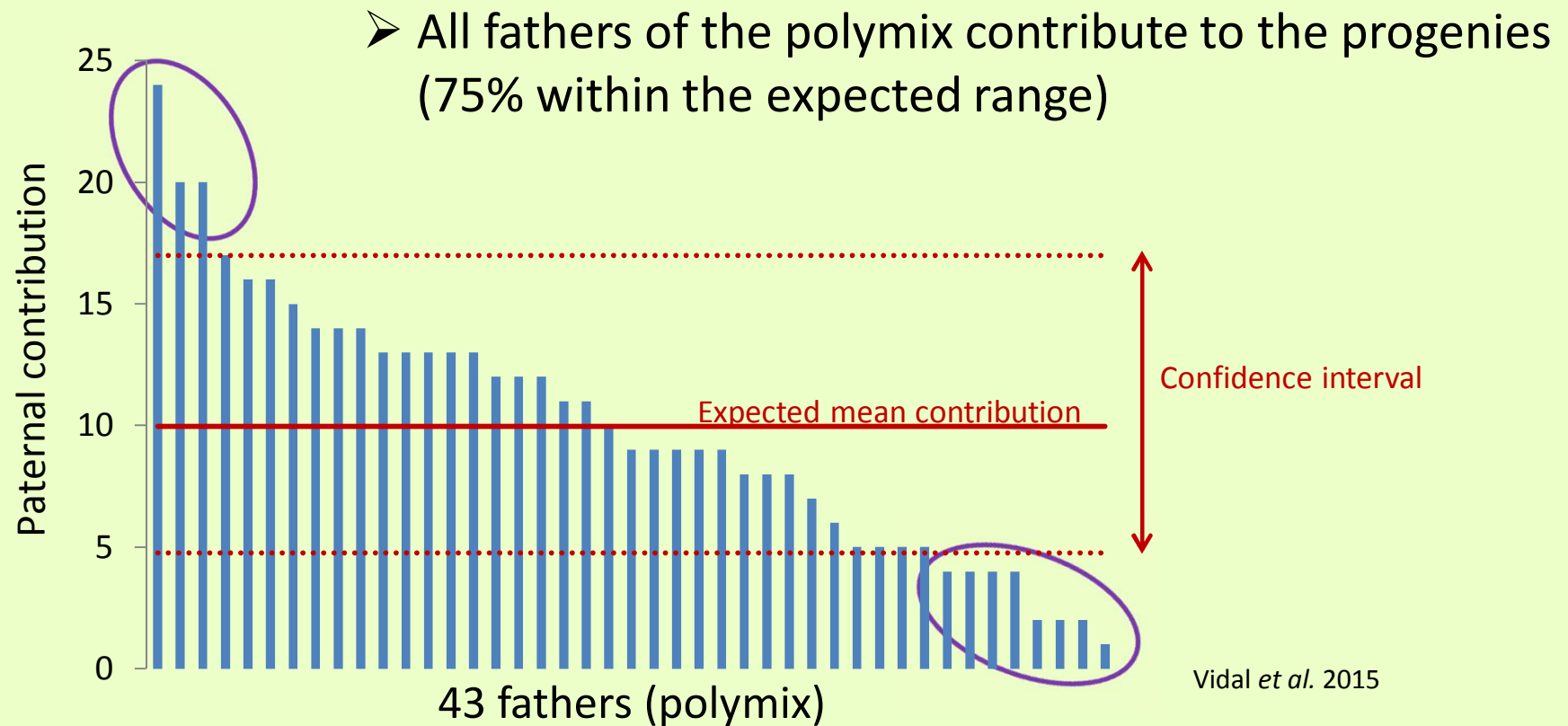
➢Efficient to maximise the number of crosses
(378 FS crosses / 428 individuals)

➢ Forward selection to establish a seed orchard

■ pre-selection (with / without diversity constraint) based on partial-pedigree information

■ final selection based on full-pedigree information

■ **Forward (pre-selection without constraint)**

■ **Forward (pre-selection with diversity constraint)**

■ **Backward**

➢ Low impact of pre-selection startegy

➢ Similar or higher genetic gains for forward selection vs. backward selection

2/ Polymix breeding with paternity recovery

- ➢ **Advantages of polymix breeding with paternity recovery:**
  - ▪ Simplification of the mating design
  - ▪ Maximization of crosses
  - ▪ Avoid pedigree mistakes
  - ▪ Accelerate deployment of genetic gains

- ➢ **Simulations will be carried out (GenTree project):**
  - ▪ Number of polymix / number of trees per family
  - ▪ Polymix composition (number of fathers, relatedness, genetic merit)
  - ▪ Cost / benefit analysis to assess economical efficiency of this strategy

3/ Genomic selection

Population 1

184 G0

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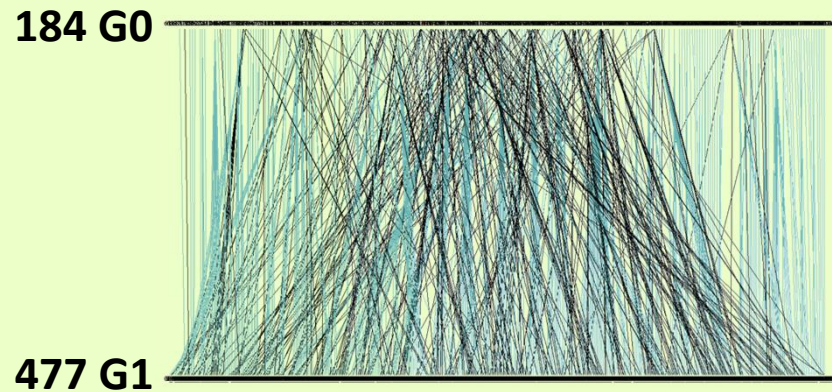
- ➢ 661 individuals (Ne 100)
- ➢ G0 + G1
- ➢ G1: 191 HS families (≈ 2.5 individuals)
- ➢ 2,500 SNPs
- ➢ Growth, sweep

Isik *et al.* 2015, Plant Science
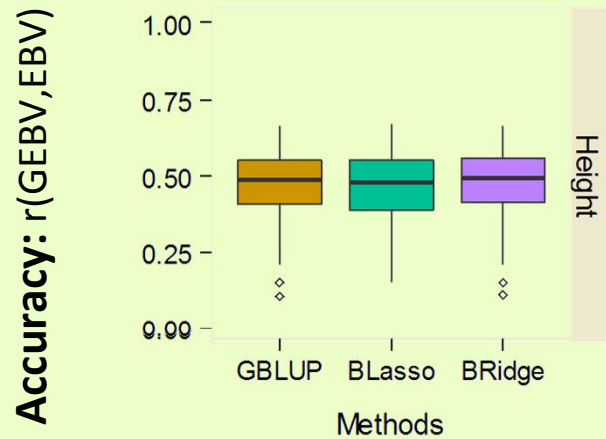
Population 2

46 G0

62 G1

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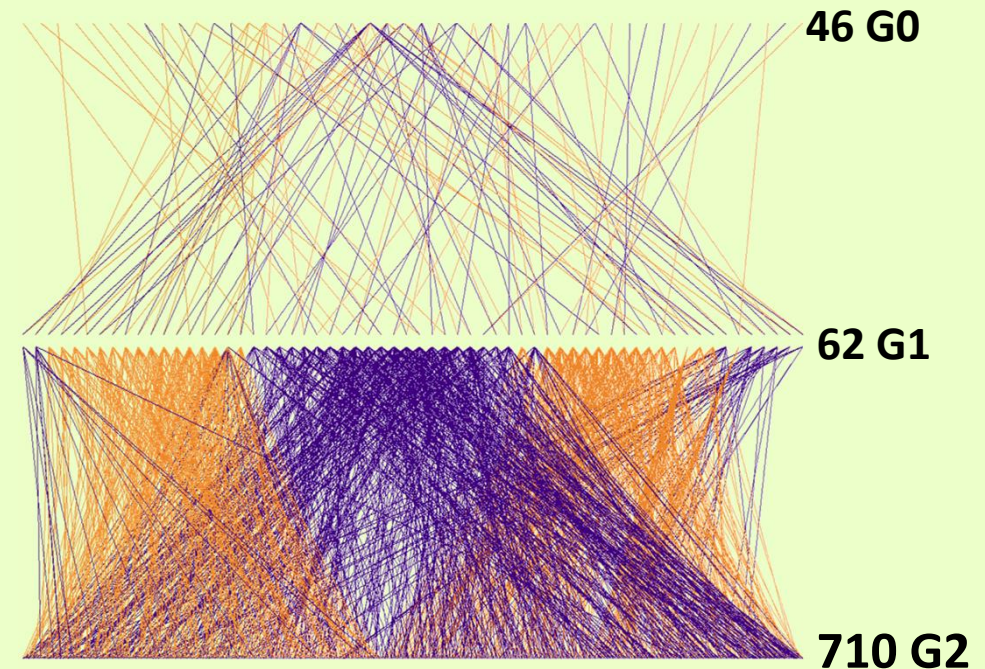
- ➢ 818 individuals (Ne=24)
- ➢ G2 + all their ancestors (G0 + G1)
- ➢ G2: 32 HS families (≈ 22 individuals)
- ➢ 4,300 SNPs
- ➢ Growth, sweep

Bartholome *et al.,* under review

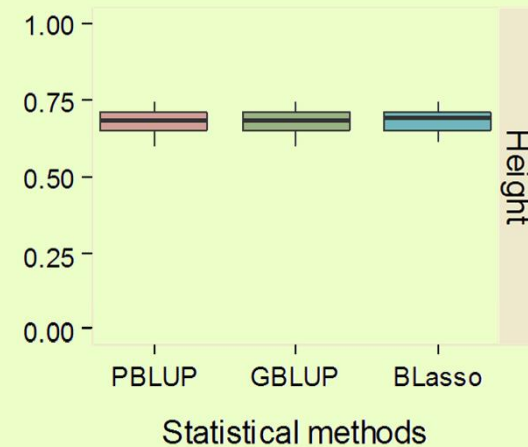# 3/ Genomic selection

Population 1

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

Population 2

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

➢ Higher accuracy in Population2

➢ Similar results whatever the method (GBLUP vs Bayesian methods)

➢ Similar accuracy for the pedigree-based method (PBLUP) and marker-based methods

Conclusion

Operational

Short term

Mid-term

1/ Identities / pedigree control
- Higher accuracy for BLUP evaluation
- First mandatory step to implement MM in the breeding program

2/ Polymix breeding with paternity recovery
- To simplify the mating process and maximize number of families evaluated
- Optimization needed with simulations

3/ Genomic selection
- Accuracy mainly due to pedigree recovery
  → Increase number of trees per FS families and number of markers to be able to predict mendelian sampling
- First step = implementation of realized genetic relationship (single-step method (Legara *et al.*, 2014) to combine pedigree and marker information)
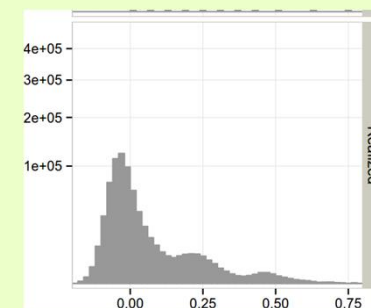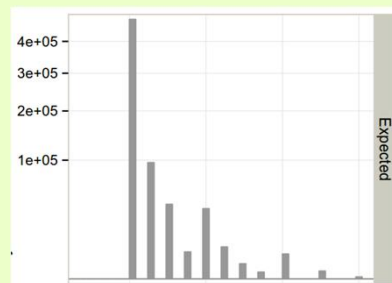
Acknowledgements

Jérôme Bartholomé, Christophe Plomion, Annie Raffin, Emilie Chancerel, Isabelle Lesur, Marc Kleinhentz

Marjorie Vidal, Luc Harvengt, Pierre Alazard

Christophe Boury, Erwan Guichoux, Adline Delcamp

Joost van Heerwaarden, Marco Bink

Fikret Isik