



HAL
open science

Soutenance - Sélection génomique chez le pin maritime

Victor Papin

► **To cite this version:**

Victor Papin. Soutenance - Sélection génomique chez le pin maritime. Amélioration des plantes. Université de bordeaux, 2023. Français. NNT: . tel-04511367

HAL Id: tel-04511367

<https://hal.inrae.fr/tel-04511367v1>

Submitted on 19 Mar 2024

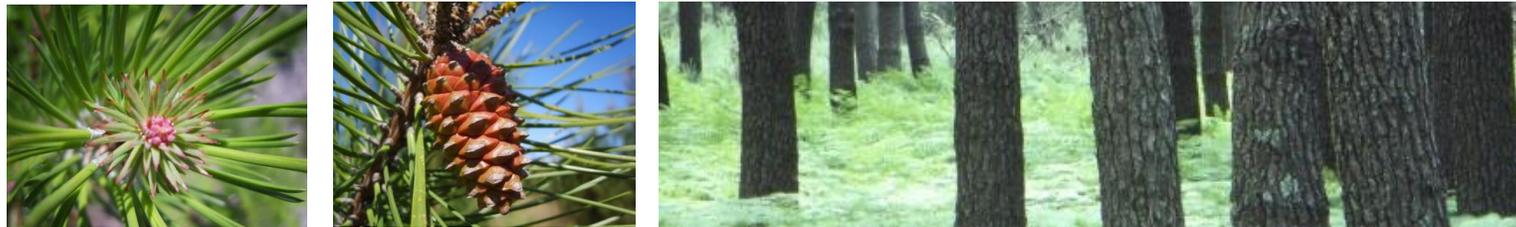
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

Soutenance de thèse - 14 décembre 2023

Sélection génomique chez le pin maritime



Victor Papin

Sous la direction de Laurent Bouffier  et Leopoldo Sanchez 

La sélection des arbres forestiers

Caractéristiques des arbres forestiers

- Espèces pérennes
- Domestication récente
- Production de bois dans les forêts de plantation
- Plantations à partir de matériel amélioré

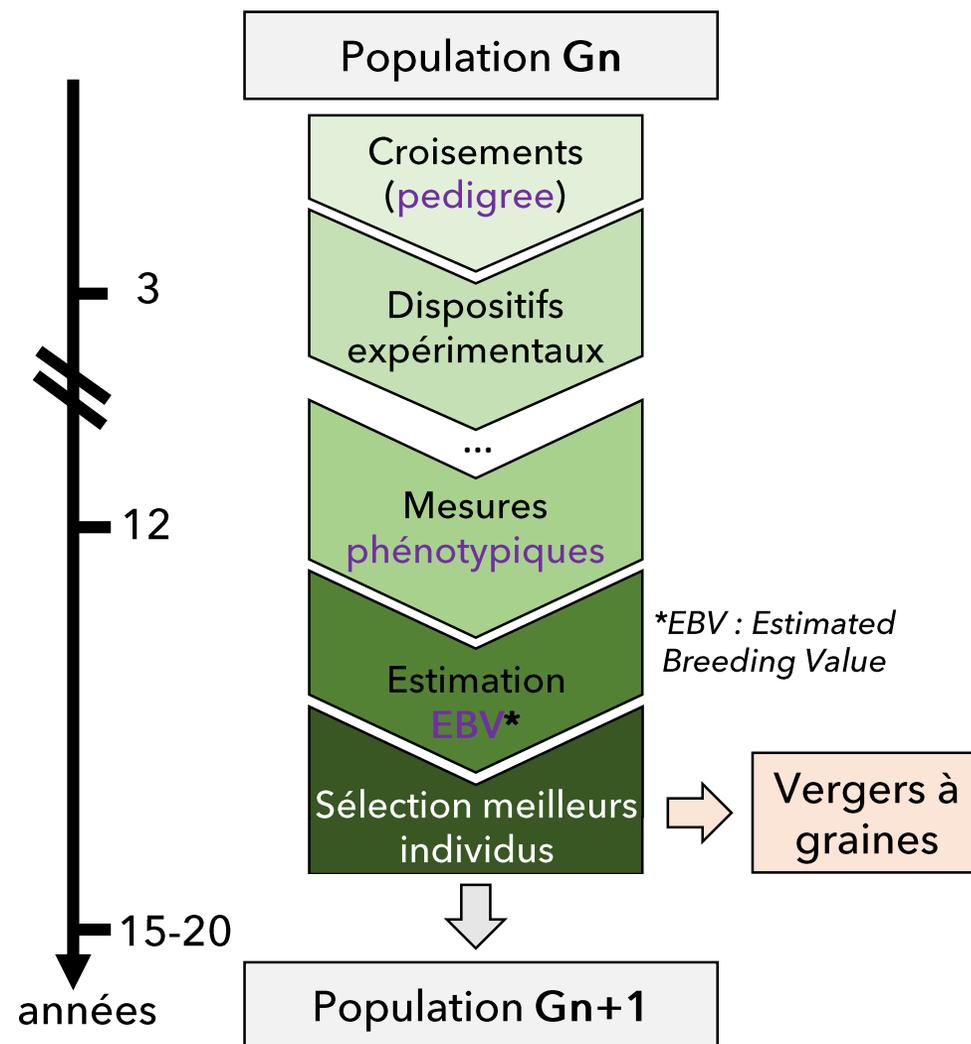
Objectifs de sélection

- Volume de bois
- Aspect qualitatif :
 - Rectitude du tronc
 - Densité du bois
 - Branchaison
- Résistance aux maladies



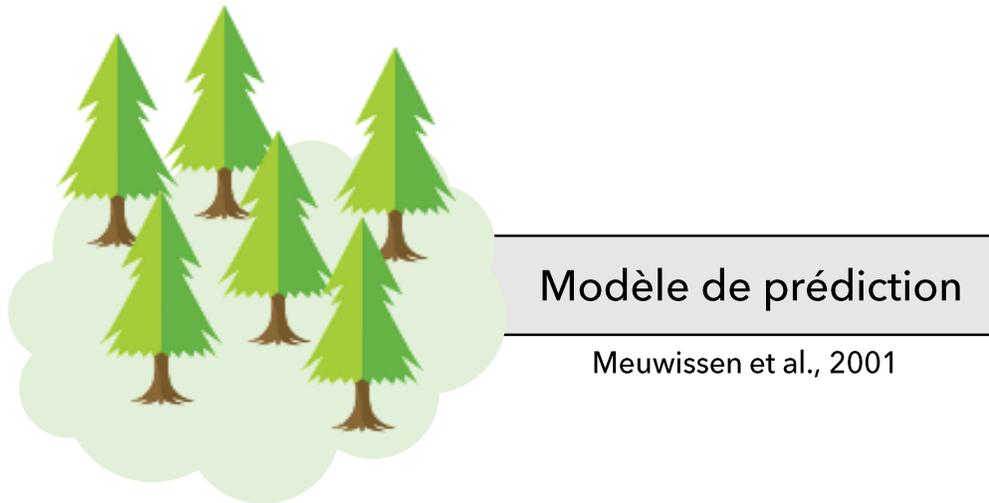
Sélection basée sur les **phénotypes** et l'**apparentement**

Schéma de sélection récurrente chez les conifères



- Demande croissante
- Diversification des usages
- Changement climatique
- Pressions biotiques

Principe et potentialités de la sélection génomique (GS)



Population de calibration

- ✓ Données phénotypiques
- ✓ Données de génotypage

Principe et potentialités de la sélection génomique (GS)



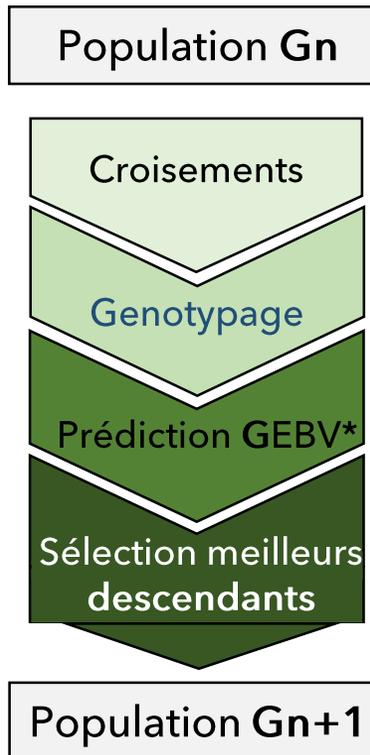
Modèle de prédiction

Meuwissen et al., 2001

Population de calibration

- ✓ Données phénotypiques
- ✓ Données de génotypage

Schéma génomique



Equation du sélectionneur

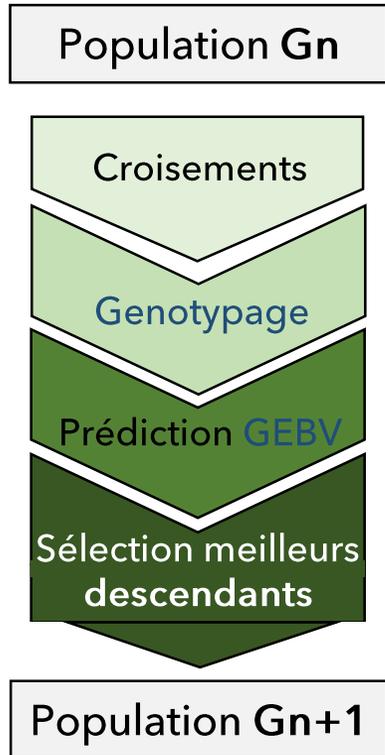
$$\Delta G = \frac{i \uparrow \times r \times \sigma_A}{T \downarrow}$$

Intensité de sélection (green arrow pointing to i)
Précision de sélection (green arrow pointing to r)
Variance génétique additive (green arrow pointing to σ_A)
Gain génétique (green arrow pointing to ΔG)
Intervalle de temps entre générations (green arrow pointing to T)

*GEBV : Genomic Estimated Breeding Value

Précision de la GS chez les arbres forestiers

Schéma génomique



Précision :

Corrélation entre GEBV prédites et GEBV vraies

Prédiction basée sur
la génomique



Moyenne parentale
& aléa de méiose
(effet mendélien)

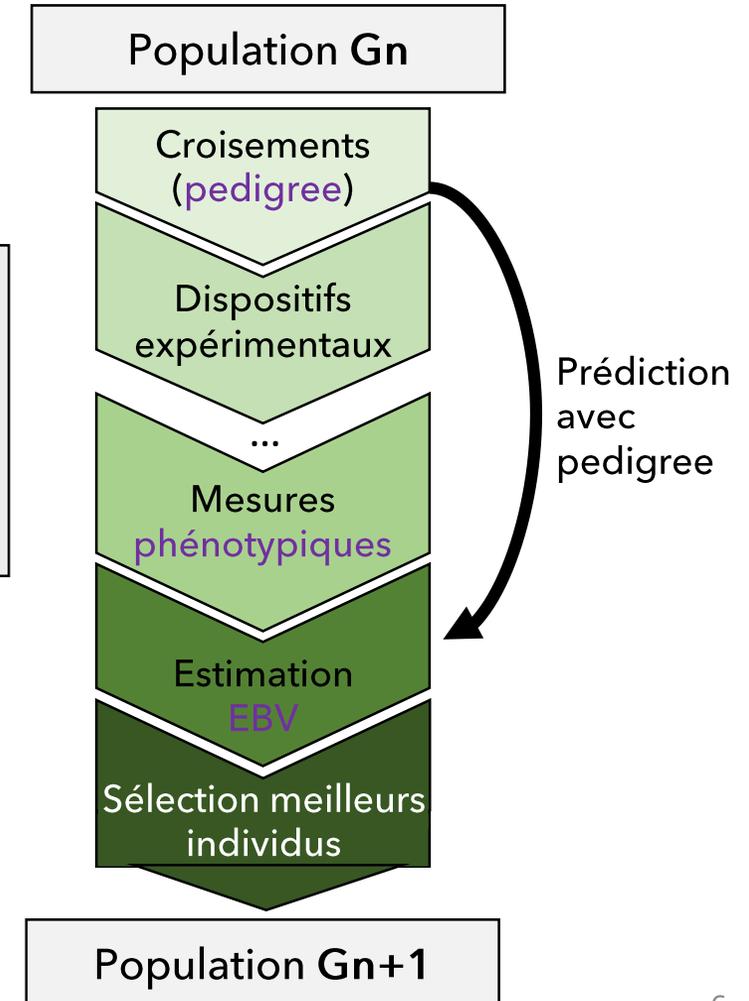
Prédiction basée sur
le pedigree



Moyenne parentale

→ Précisions équivalentes

Schéma conventionnel



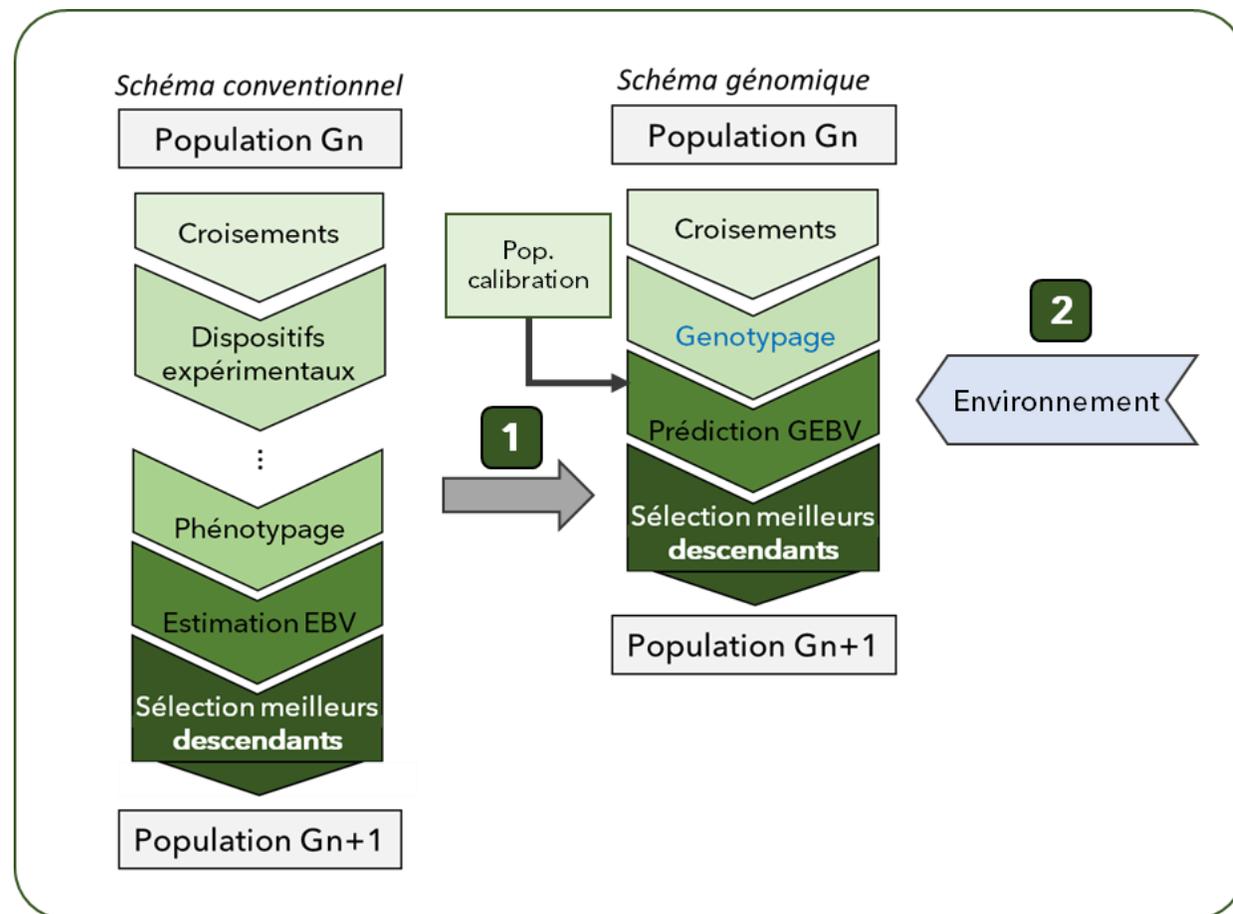
L'information environnementale en sélection

- Peu de considération pour l'information environnementale en sélection génomique
- Dispositifs d'évaluation multi-sites coûteux à mettre en place et difficiles à maintenir dans le temps
- Changements climatiques forts attendus



→ La GS pour faciliter l'intégration de l'information environnementale en sélection

Questions scientifiques



1

Quelle est la précision de la GS en intra-famille chez les arbres forestiers ?

2

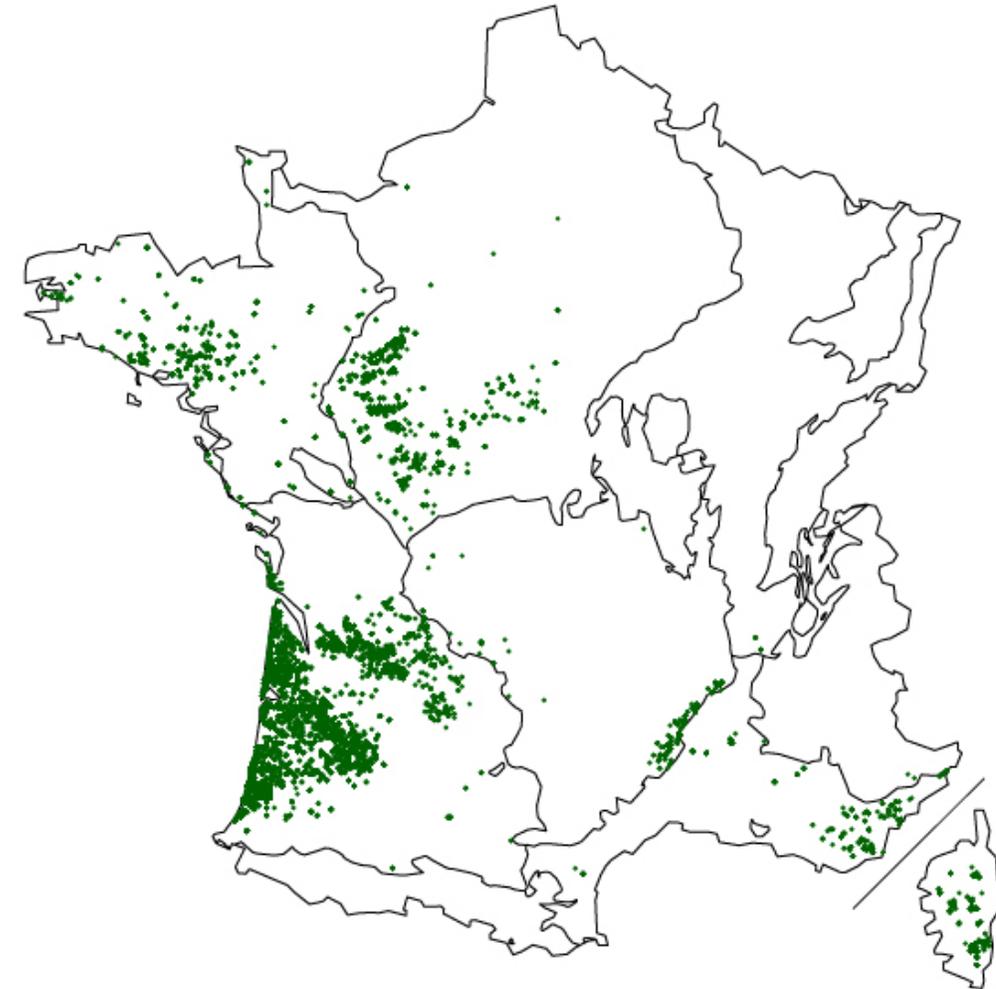
Dans quelles conditions la GS démontre-t-elle un avantage prédictif par rapport à la sélection conventionnelle ?

2

Comment intégrer l'information environnementale dans la sélection des arbres forestiers ?

Objet d'étude : le pin maritime

- Le pin maritime (*pinus pinaster* Ait) est l'essence la plus plantée en France : **740 000 ha** en peuplement pur (IGN)
- Programme de sélection récurrente débuté dans les années 1960 : diffusion de variétés améliorées en terme de **croissance** et de **rectitude**



● Présence du pin maritime dans les grandes régions écologiques

Source : inventaire-forestier.ign.fr

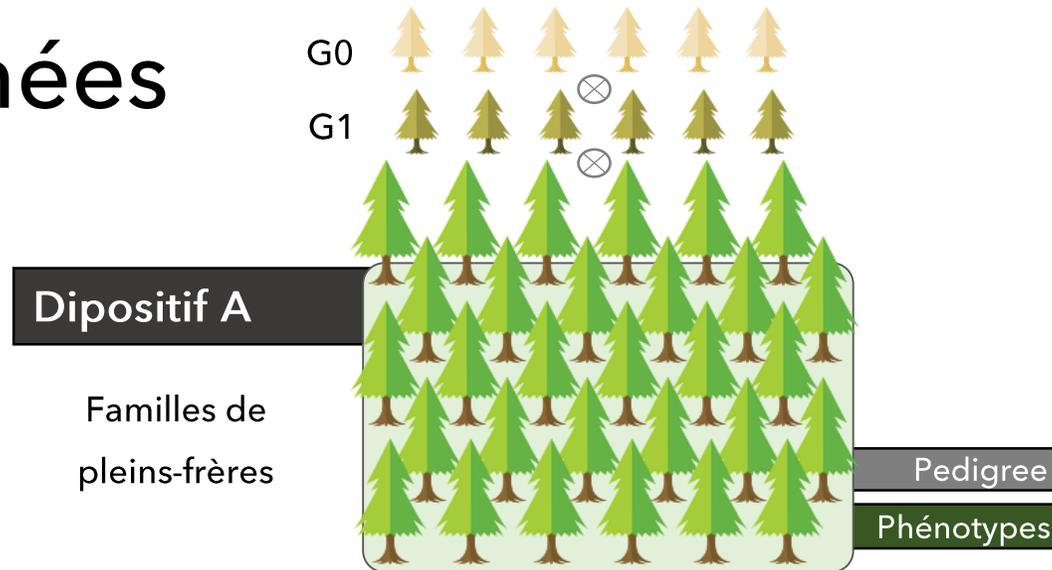
A photograph of a dense forest of tall, thin trees, likely pines or spruces, with a lush green fern forest floor. The trees are arranged in a regular pattern, creating a sense of depth and perspective. The lighting is soft, suggesting an overcast day or a shaded forest interior.

Précision de la prédiction génomique

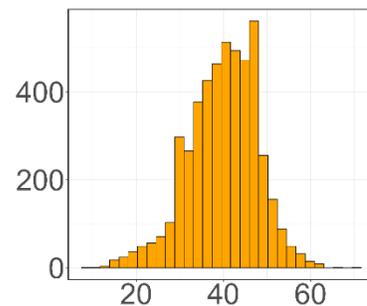
Précision intrafamiliale

Comparaison avec la sélection
conventionnelle basée sur le pedigree

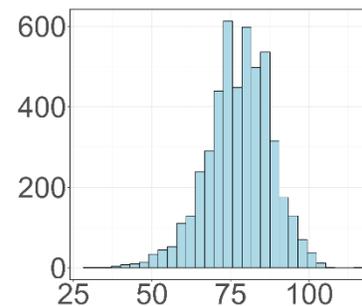
Design et données



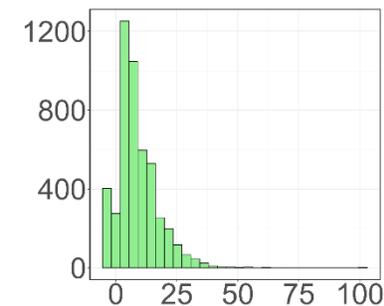
Données phénotypiques



Circonférence
(cm) - à 12 ans

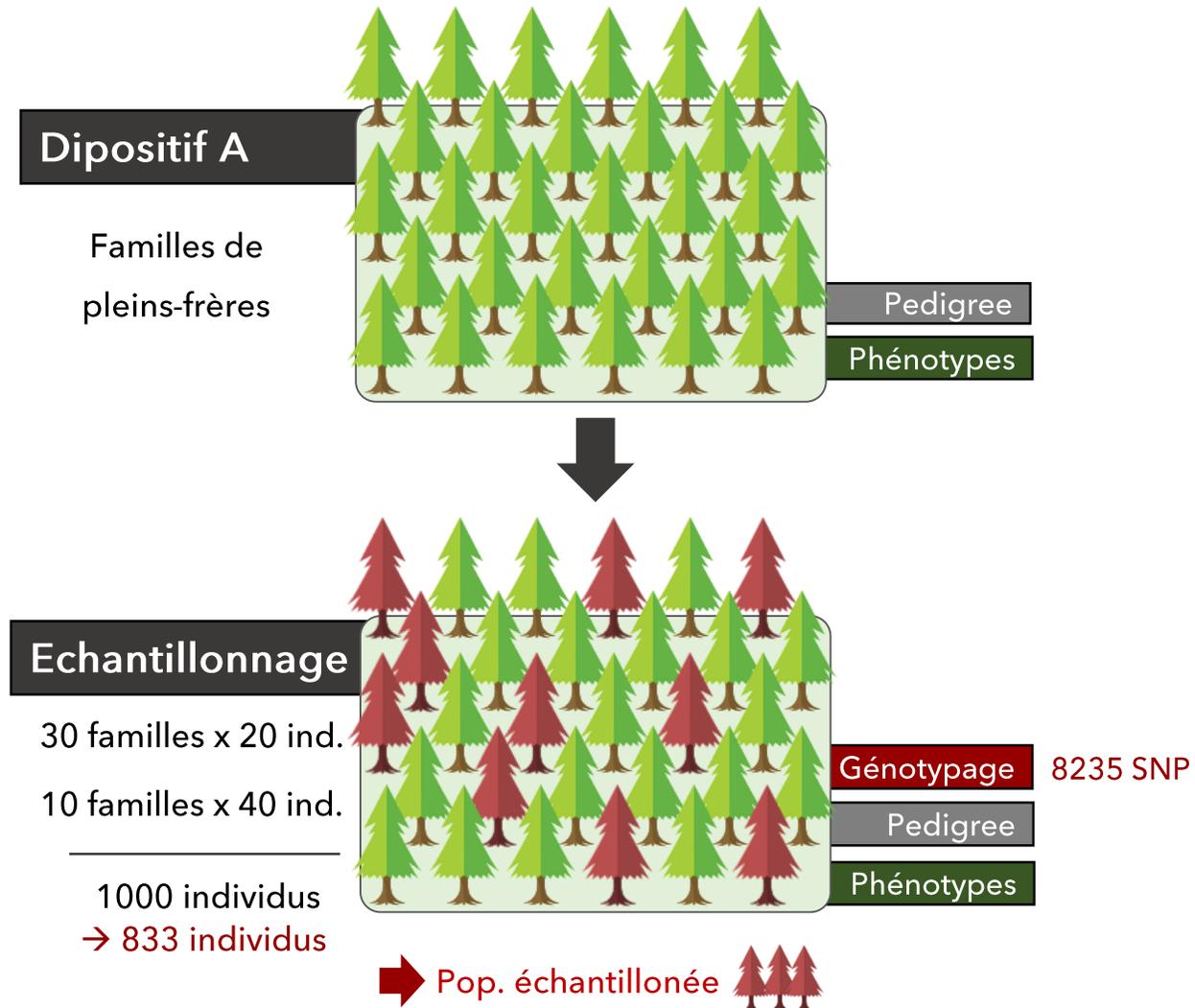


Hauteur
(dm) - à 12 ans

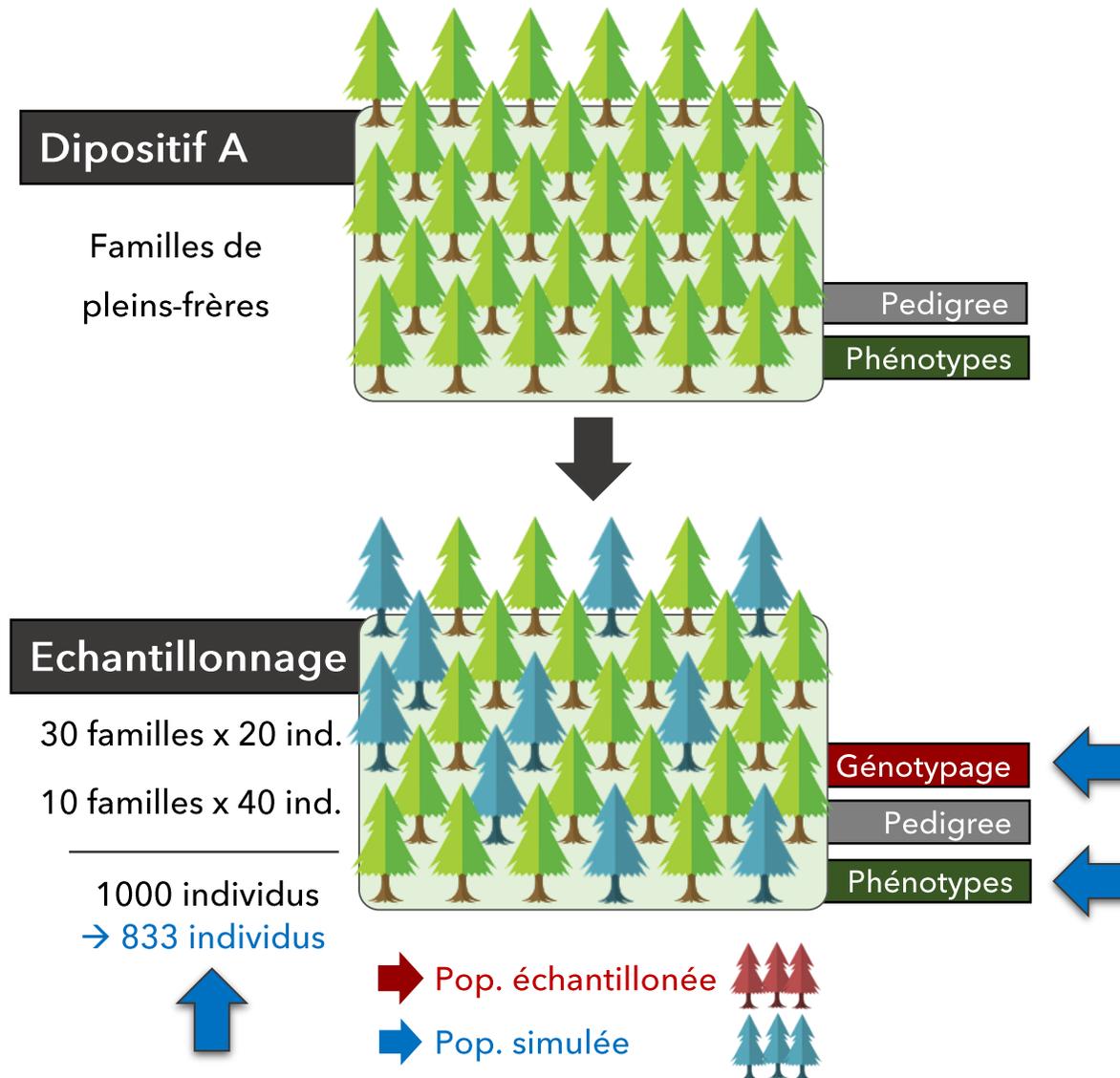


Ecart verticalité
(cm) - à 8 ans

Échantillonnage



Échantillonnage



Modèle de GS et cross-validation

- **Modèle** : $y = 1\mu + Zu + \epsilon$

- $u \sim N(0, \mathbf{A}\sigma_u^2) \rightarrow \mathbf{ABLUP}$

- $u \sim N(0, \mathbf{G}\sigma_u^2) \rightarrow \mathbf{GBLUP}$

- **Précision de prédiction** :

- Déterminée par cross-validation

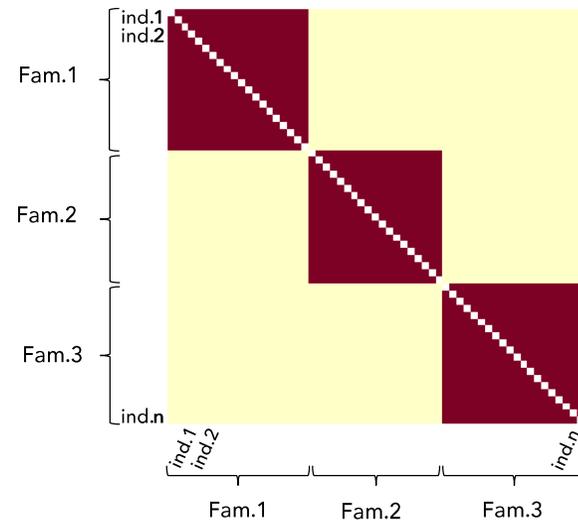
$\rightarrow \text{corr}(\text{GEBV}_{\text{prédits}}, \text{Pheno})$

- Deux échelles :

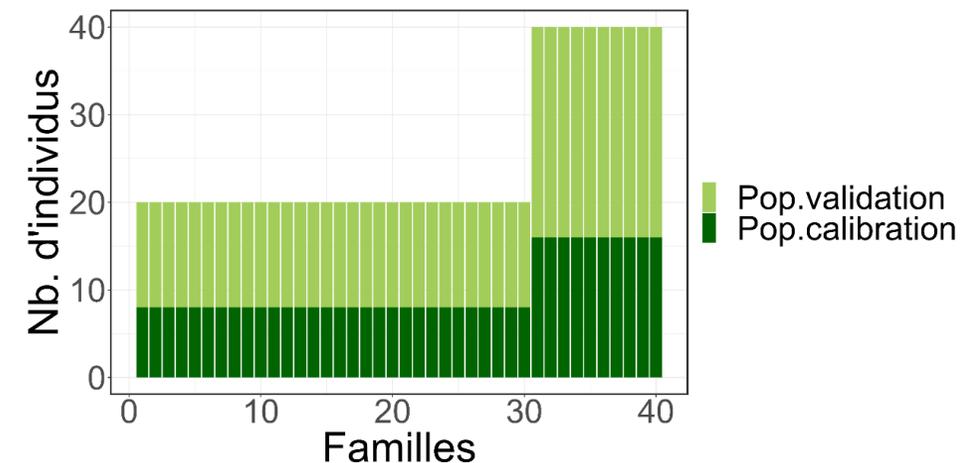
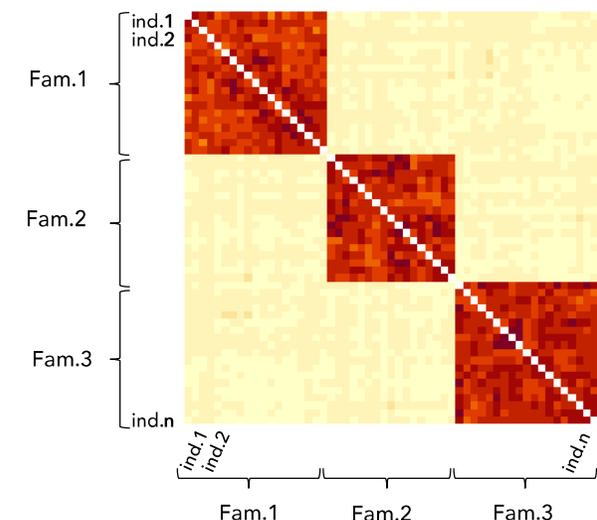
\rightarrow Précision globale

\rightarrow Précision intrafamiliale

Matrice A
Apparement pedigree



Matrice G
Apparement moléculaire



Modèle de GS et cross-validation

- **Modèle** : $y = 1\mu + Zu + \epsilon$

- $u \sim N(0, \mathbf{A}\sigma_u^2) \rightarrow \mathbf{ABLUP}$

- $u \sim N(0, \mathbf{G}\sigma_u^2) \rightarrow \mathbf{GBLUP}$

- **Précision de prédiction** :

- Déterminée par cross-validation

- $\rightarrow \text{corr}(\text{GEBV}_{\text{prédits}}, \text{Pheno})$

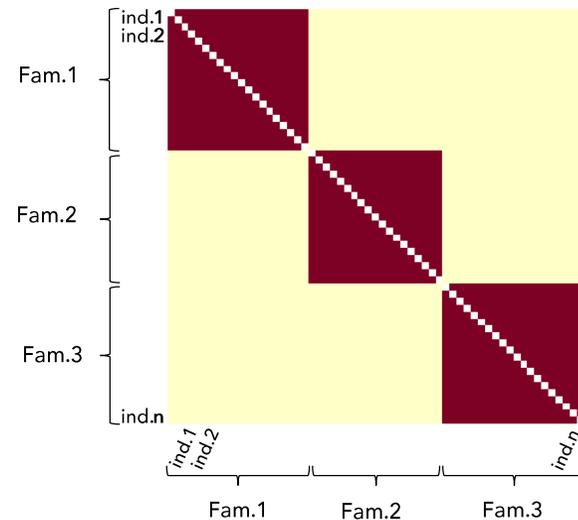
- Deux échelles :

- \rightarrow Précision globale

- \rightarrow Précision intrafamiliale

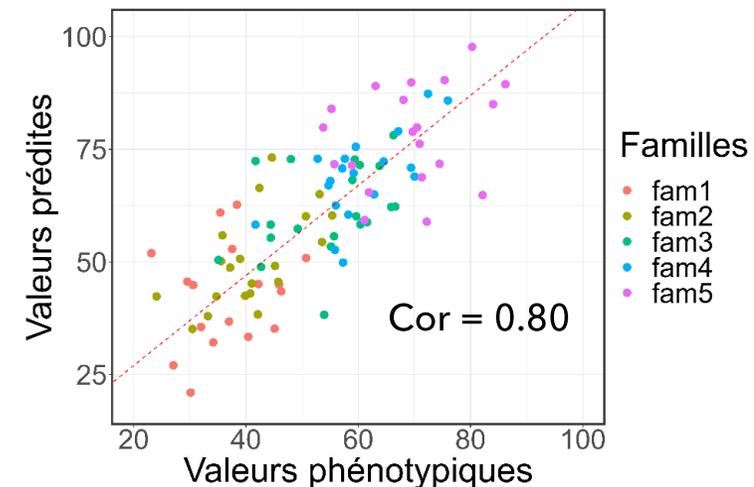
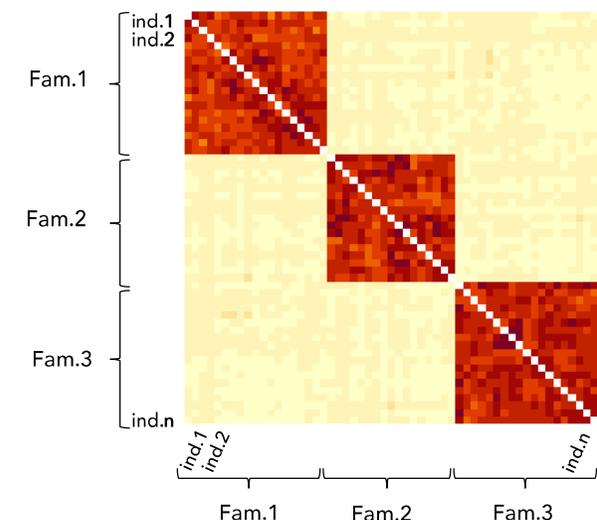
Matrice A

Apparemment pedigree



Matrice G

Apparemment moléculaire



Modèle de GS et cross-validation

- **Modèle** : $y = 1\mu + Zu + \epsilon$

- $u \sim N(0, \mathbf{A}\sigma_u^2) \rightarrow \mathbf{ABLUP}$

- $u \sim N(0, \mathbf{G}\sigma_u^2) \rightarrow \mathbf{GBLUP}$

- **Précision de prédiction** :

- Déterminée par cross-validation

$\rightarrow \text{corr}(\text{GEBV}_{\text{prédits}}, \text{Pheno})$

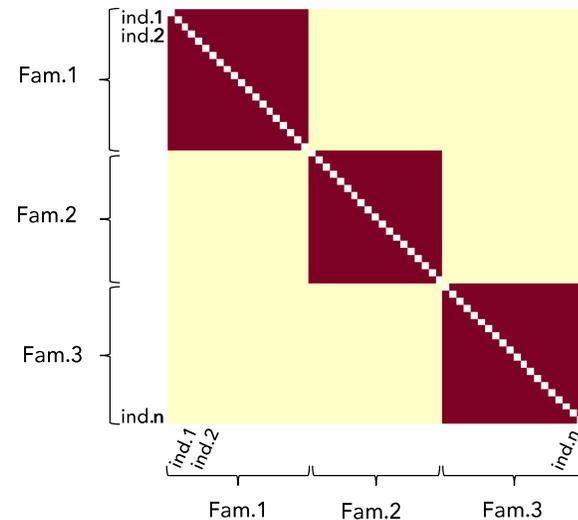
- Deux échelles :

\rightarrow Précision globale

\rightarrow Précision intrafamiliale

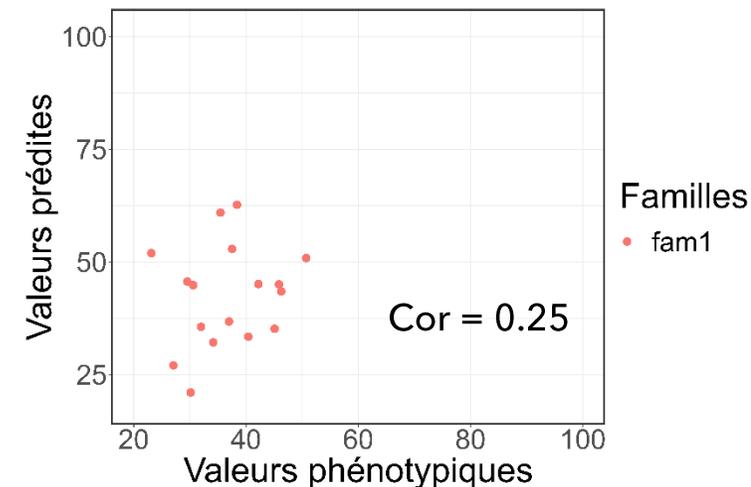
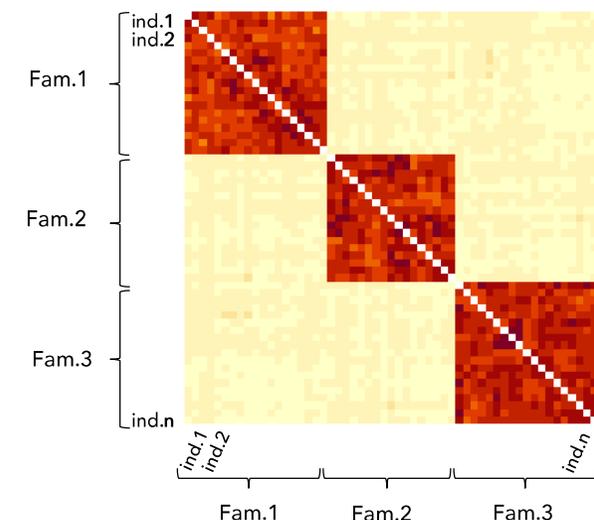
Matrice A

Apparemment pedigree

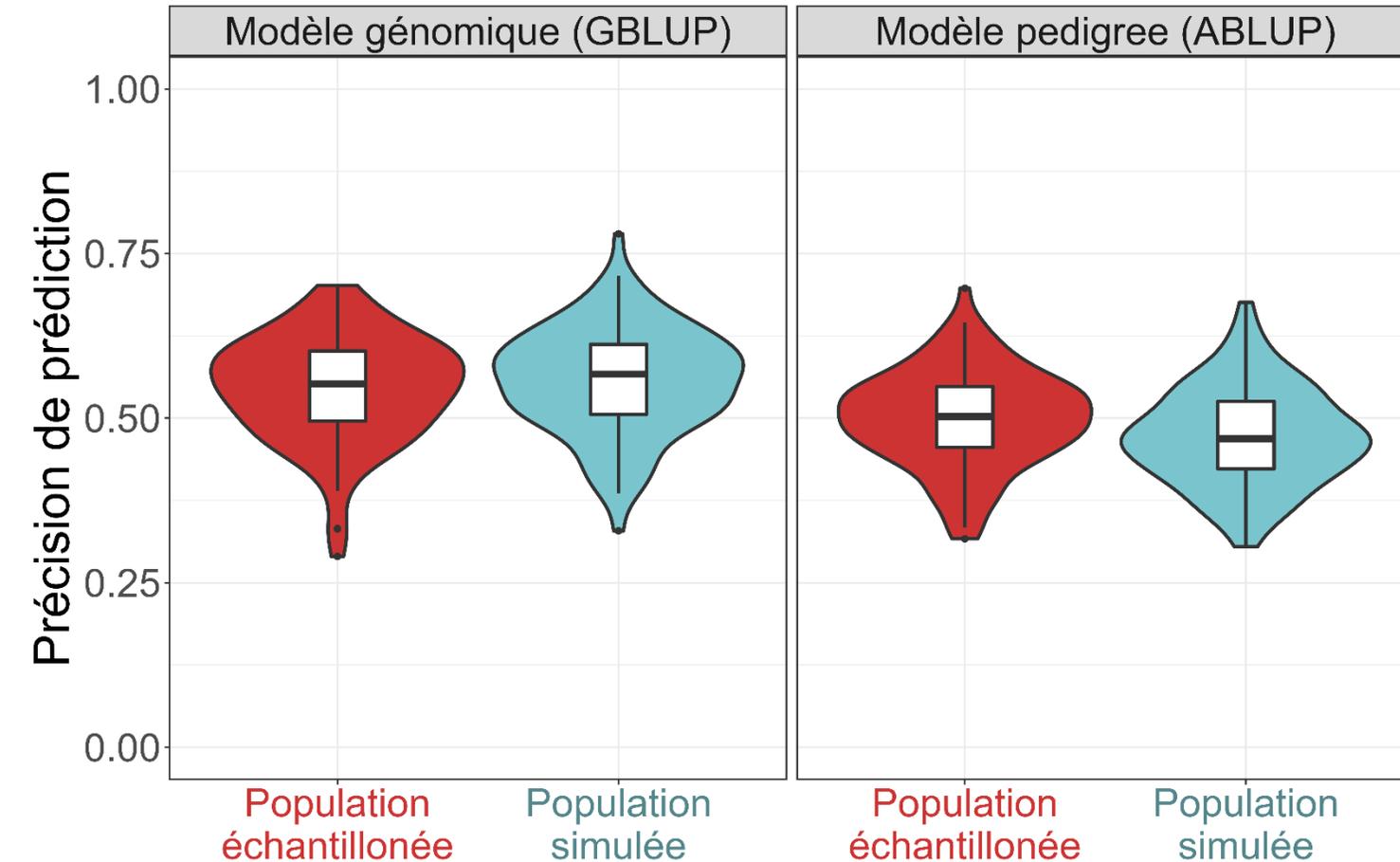


Matrice G

Apparemment moléculaire

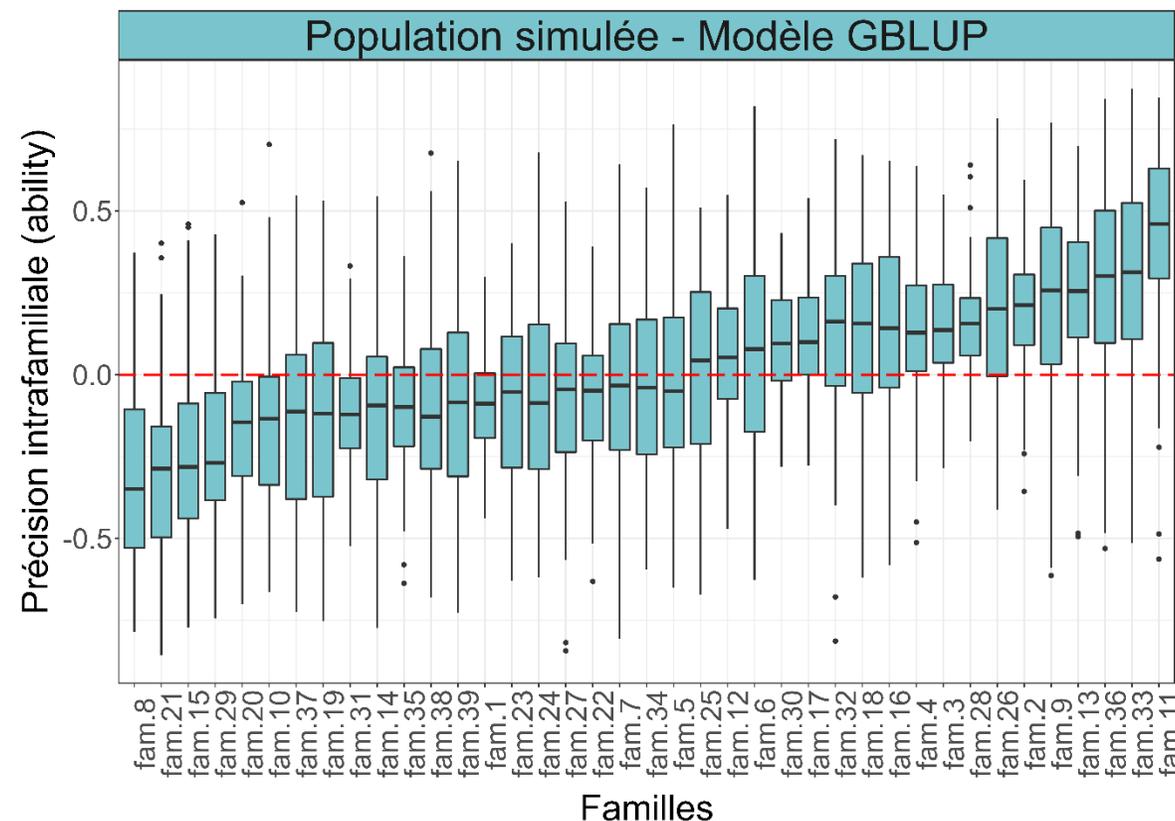
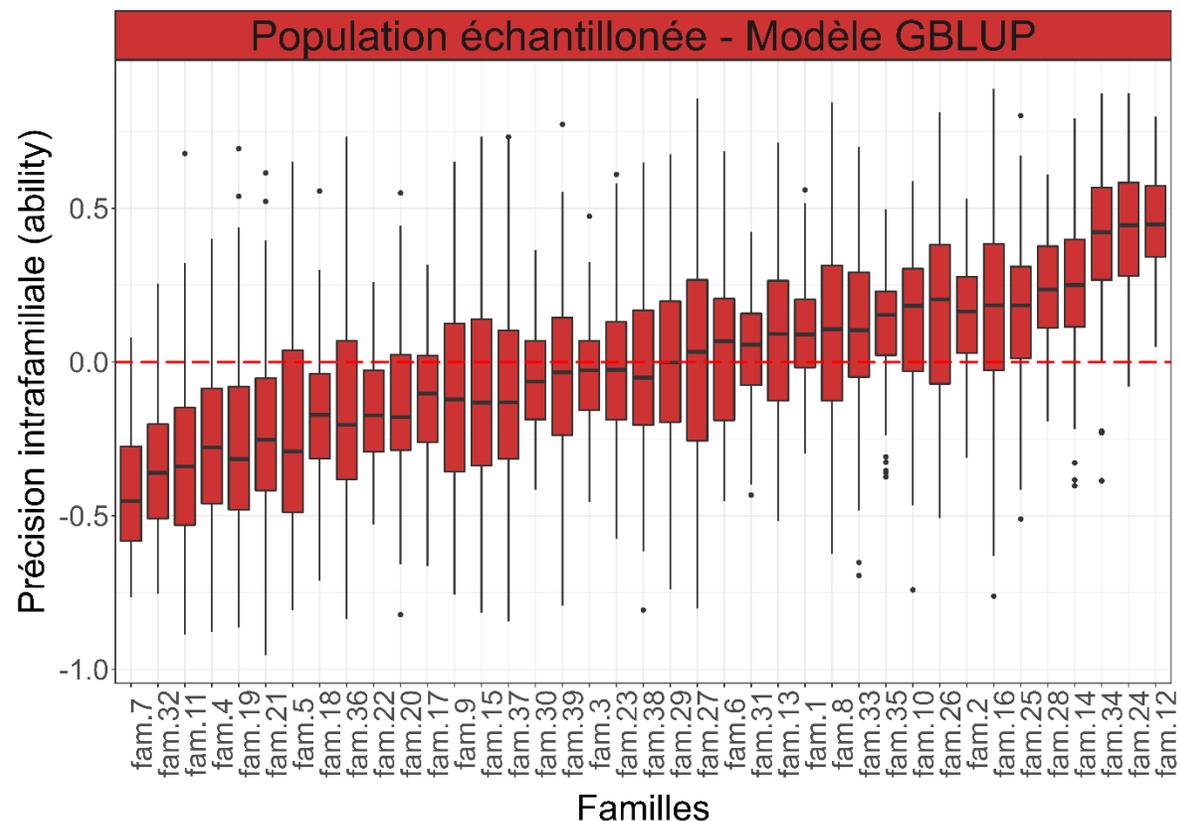


Précision globale de la GS



- Niveau de précision modéré du GBLUP, commun en génétique forestière
- Léger avantage du modèle GBLUP (0,56) par rapport au modèle ABLUP (0,50)
- Cohérence entre les résultats obtenus avec les données réelles et les données simulées

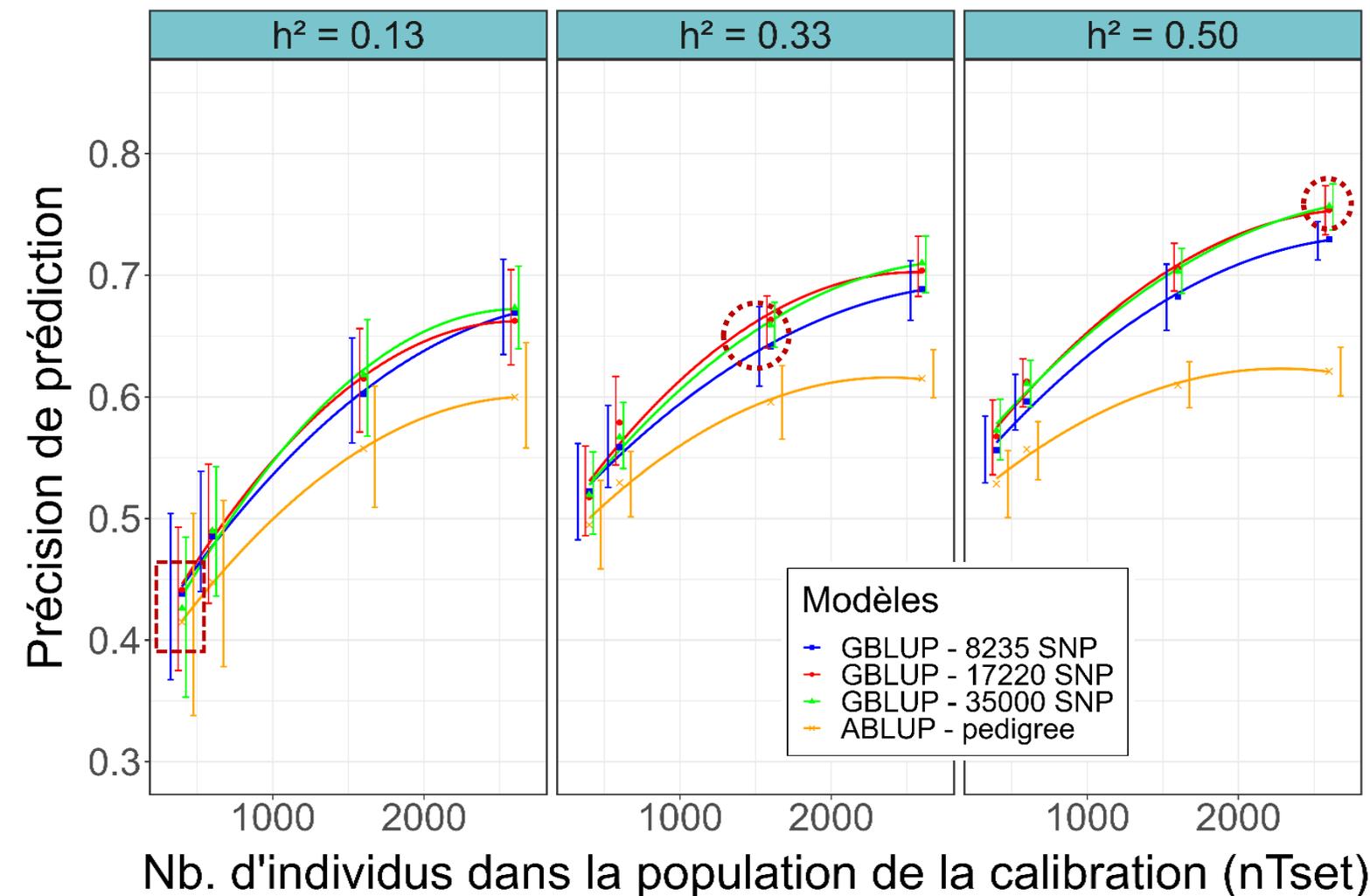
Précision intrafamiliale de la GS



- Précision très hétérogène selon les familles, nulle en moyenne
- Cohérence entre les résultats obtenus avec les données réelles et les données simulées

Exploration de la précision globale de la GS

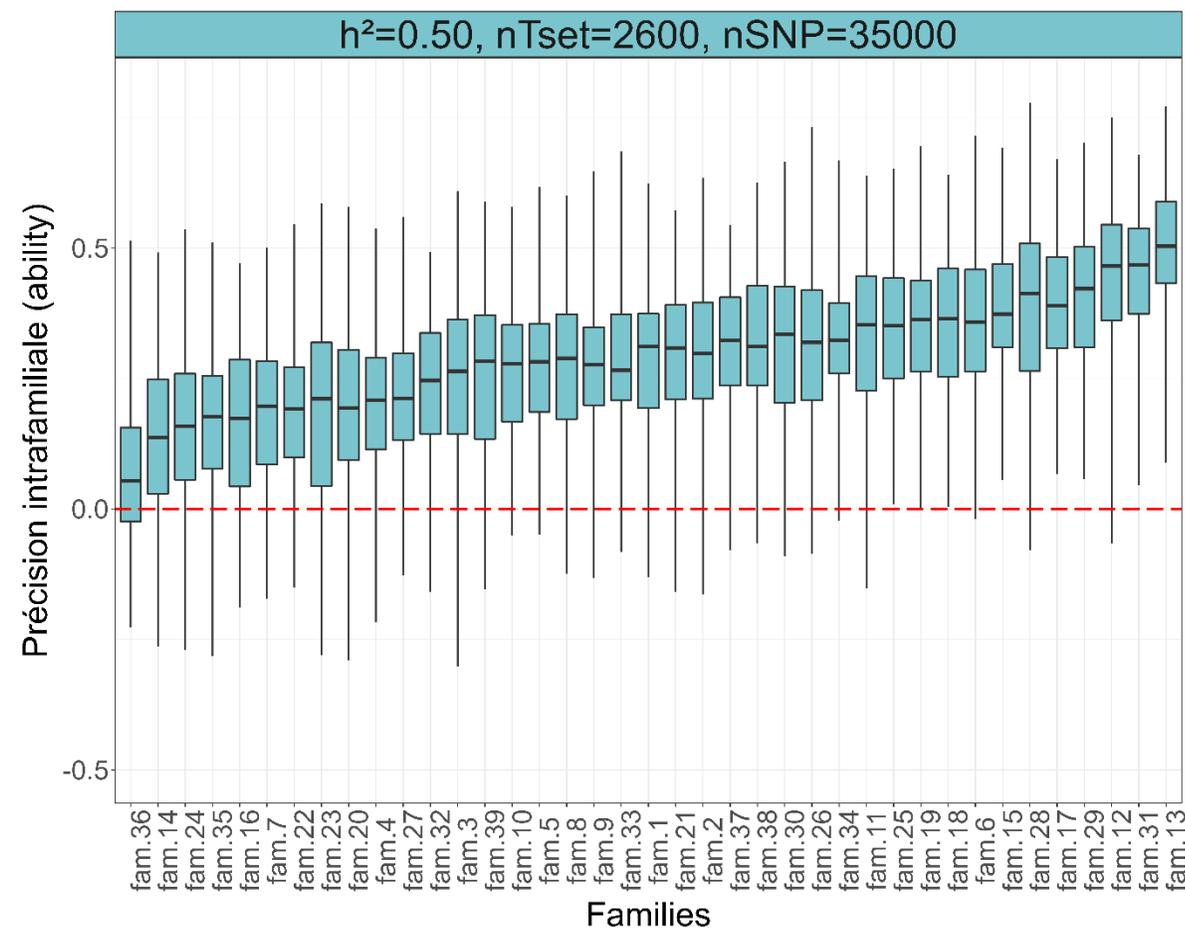
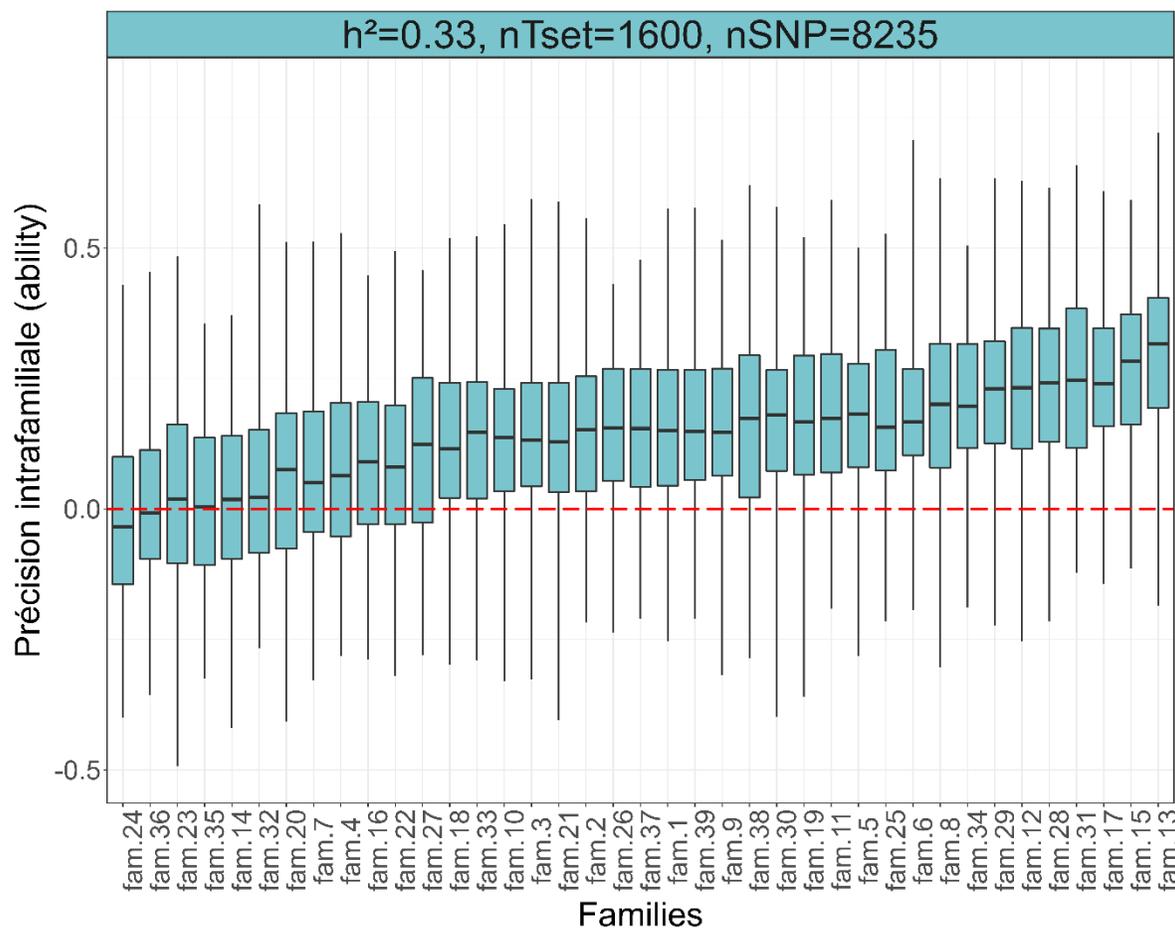
par simulation ($nTset$, $nSNP$, h^2)



- Précision globale en fonction de :
 - $nTset$: nombre d'individus inclus dans la calibration
 - $nSNP$: nombre de SNP caractérisant chaque individu
 - h^2 : héritabilité du trait simulé
- Précision de prédiction globale très dépendante de $nTset$
- Effet marginal de $nSNP$

Exploration de la précision intrafamiliale de la GS

par simulation ($nTset$, $nSNP$, h^2)



- Avantage du modèle GBLUP coincide avec une précision intrafamiliale positive

Facteurs explicatifs de la précision intrafamiliale de la GS

par simulation

Corrélation avec la précision de
prédiction moyenne par famille

App.calib : apparentement génomique
moyen à la pop. de calibration

Mean.G : moyenne des apparentements
génomiques

Var.G : variance des apparentements
génomiques

Het : Taux moyen d'hétérozygotie

App.G : Coefficient d'apparentement
génomique entre les deux parents

Het.P1P2 : Taux d'hétérozygotie
moyen parental

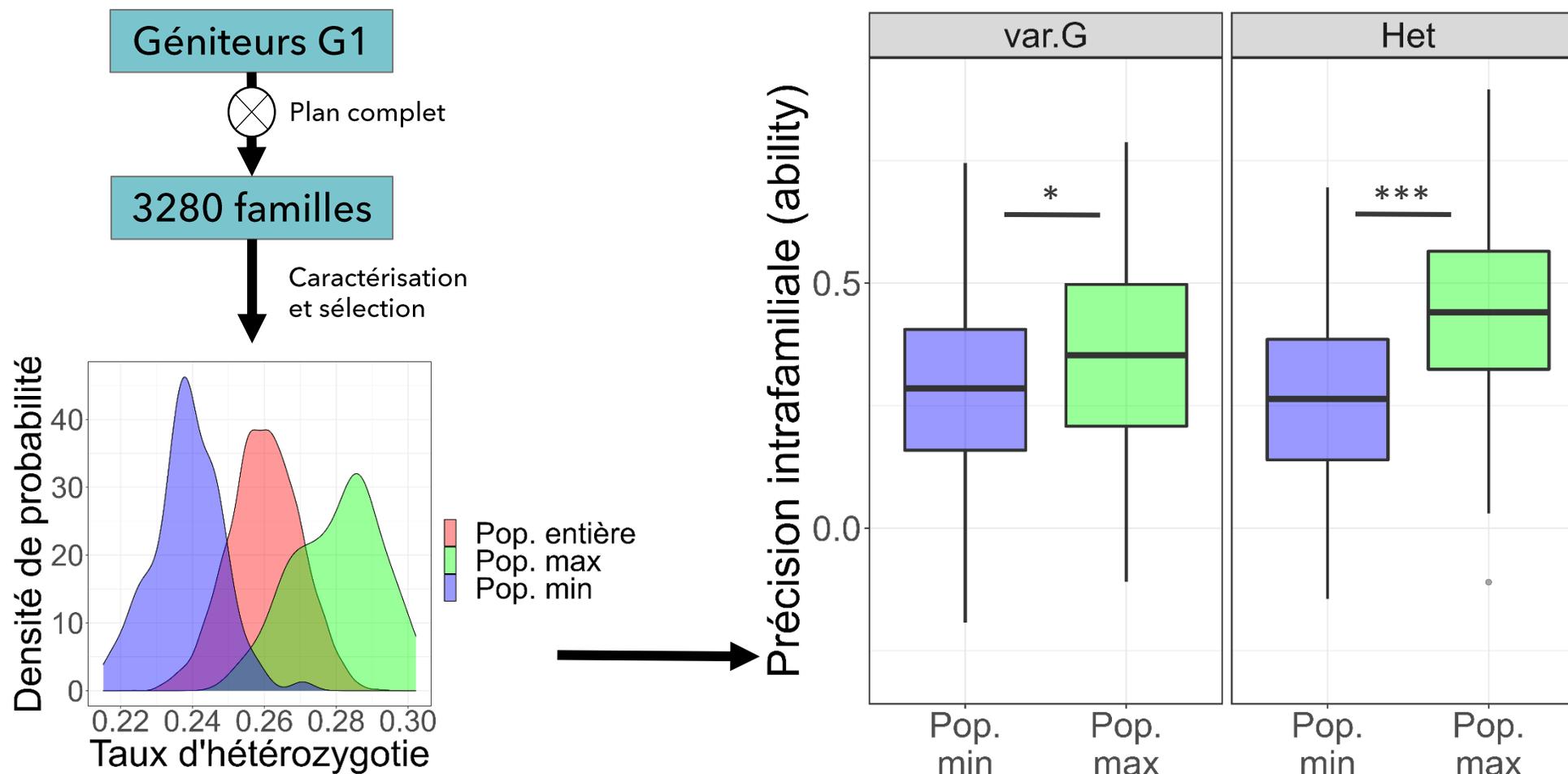
0,20 [*]	0,19 [*]
-0,19	-0,15
0,15	0,27 [*]
0,04	0,53 ^{**}
-0,01	-0,22
-0,15	0,30 [*]

Pondération des loci par leur effet

* : 5 e-2
** : 5 e-3

Facteurs explicatifs de la précision intrafamiliale de la GS

par simulation



Conclusion partie n°1

Sélection génomique dans notre population initiale

- Faible avantage du modèle génomique (GBLUP) en terme de précision de prédiction
- Approche originale : distinguer précision globale et intrafamiliale
- Précision génomique intrafamiliale très hétérogène mais en nulle en moyenne

Facteurs clés affectant la précision de la prédiction génomique

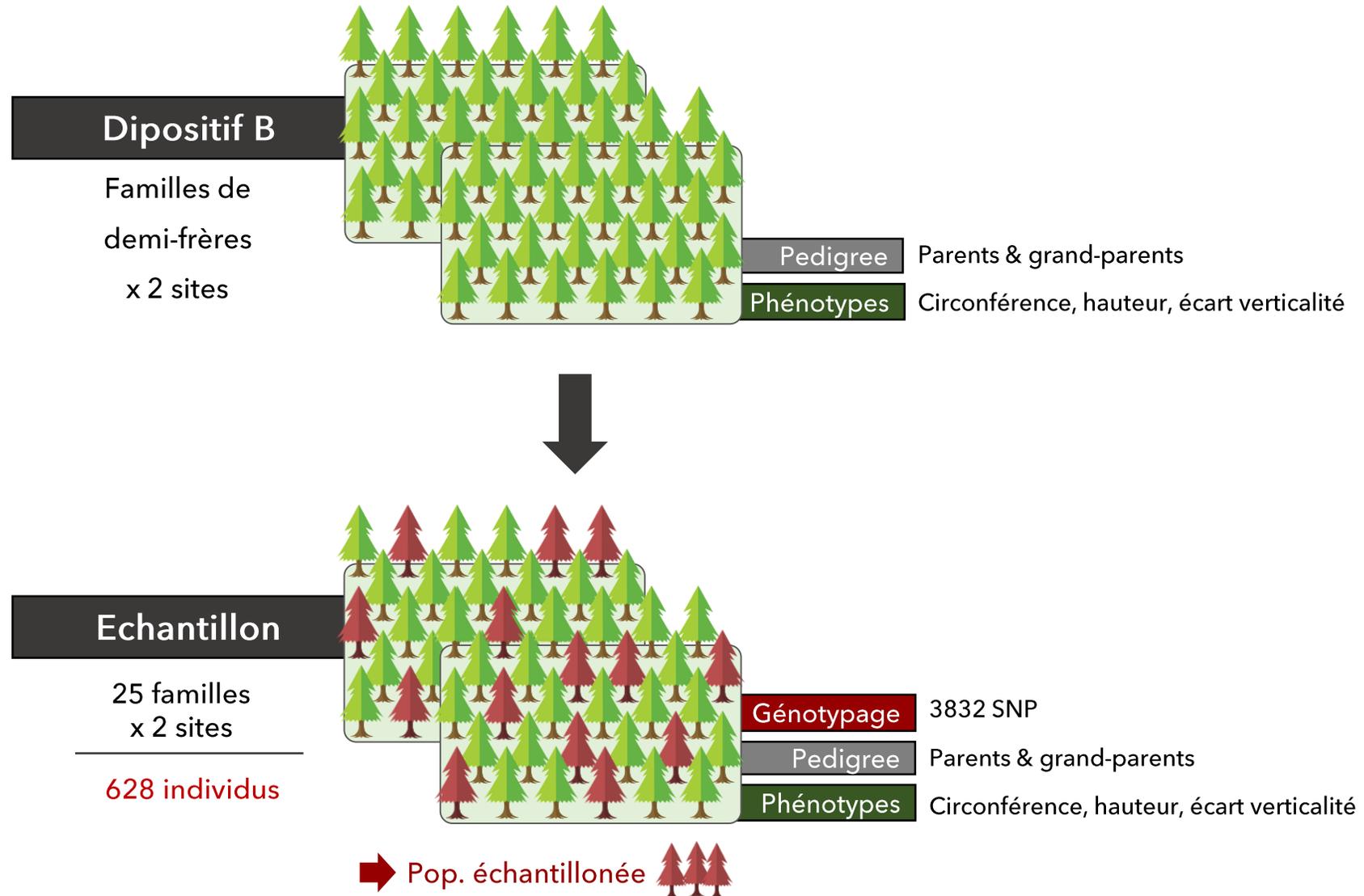
- Taille de la population de calibration : principal facteur limitant de la précision génomique
- L'avantage du modèle génomique coïncide avec une précision intrafamiliale positive
- Les familles les mieux prédites sont celles qui présentent le plus de variabilité génétique

A photograph of a dense forest with tall, thin trees and a green fern forest floor. The text is overlaid on the left side of the image.

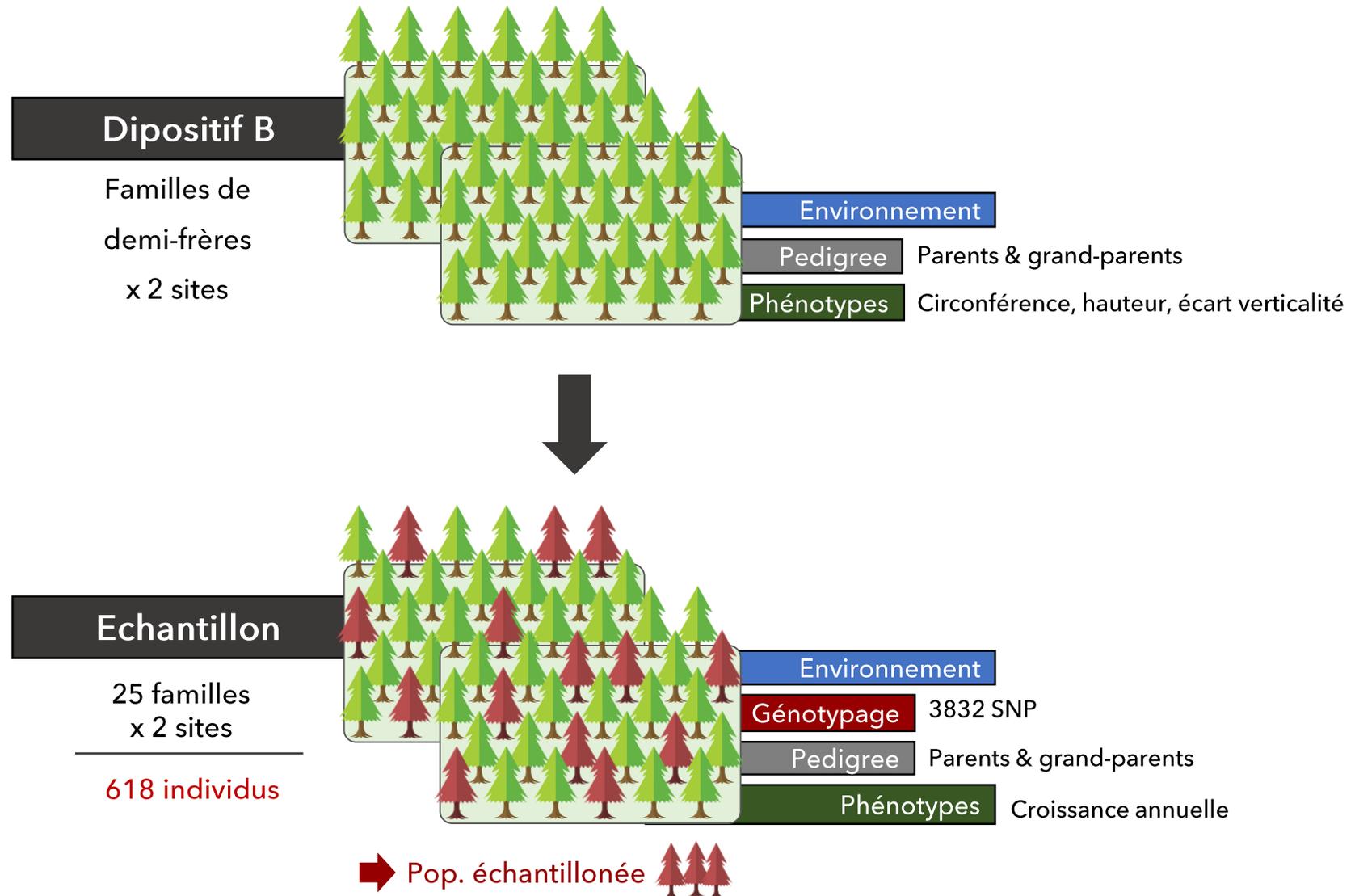
Normes de réaction génomiques

Intégration de l'information
environnementale en sélection

Design et données



Design et données



Caractérisation de la croissance annuelle

1 Prélèvement carotte de bois



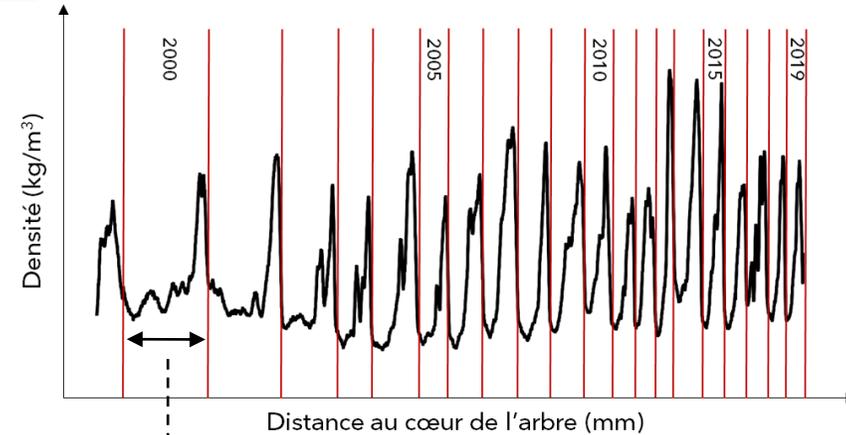
2 Découpage



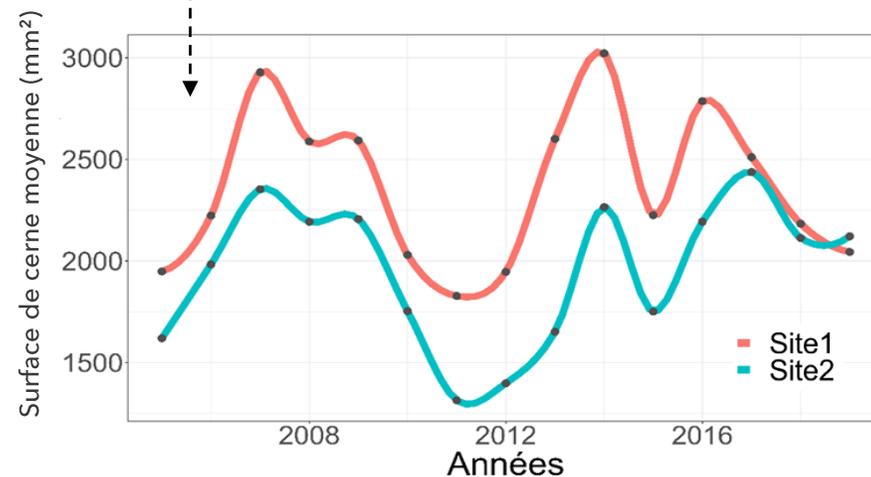
3 Radiographie aux rayons X



4 Profil de densitométrie



5 Croissance annuelle



Données environnementales

Indice d'aridité de Martonne



p : cumul des précipitations (mm)

t : température moyenne (°C) /mois



$$I = \frac{12p}{t + 10}$$

(de Martonne, 1926)



mars
-
septembre

Indice d'aridité annuel
(2005-2019)

Potentiel de croissance



Données climatiques /heure



Caractérisation du sol



Pratiques sylvicoles



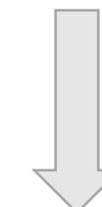
Modèle GO+

(Moreaux et al., 2020)

- Température du microclimat
- Potentiel hydrique du tronc

→ Potentiel de croissance journalier

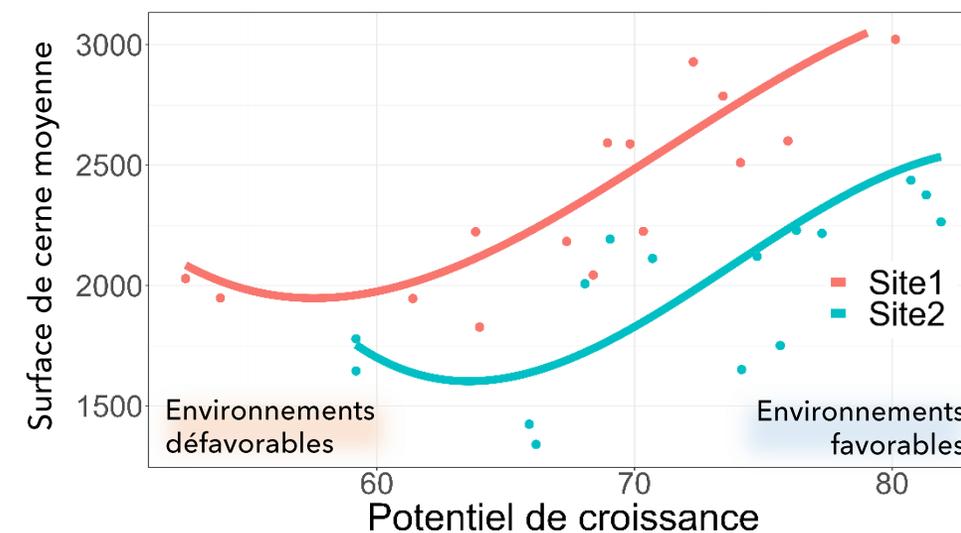
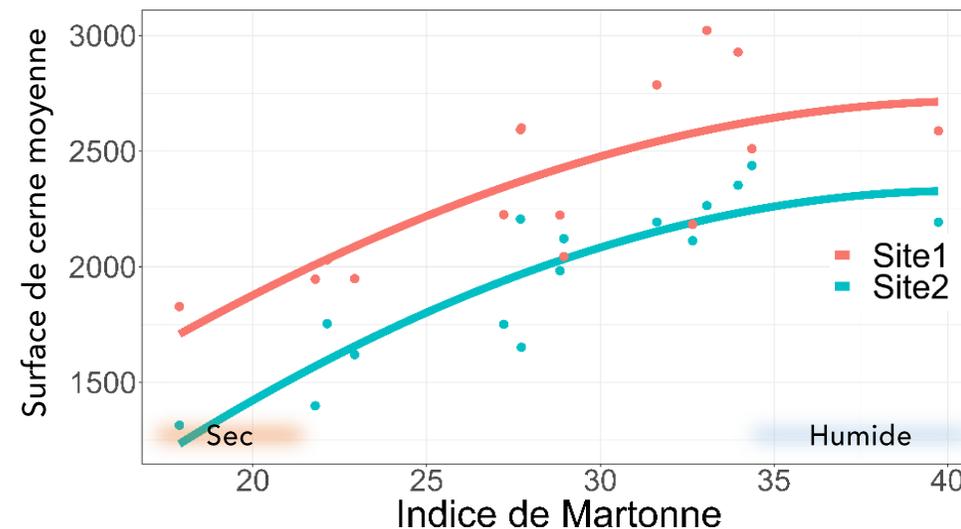
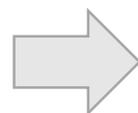
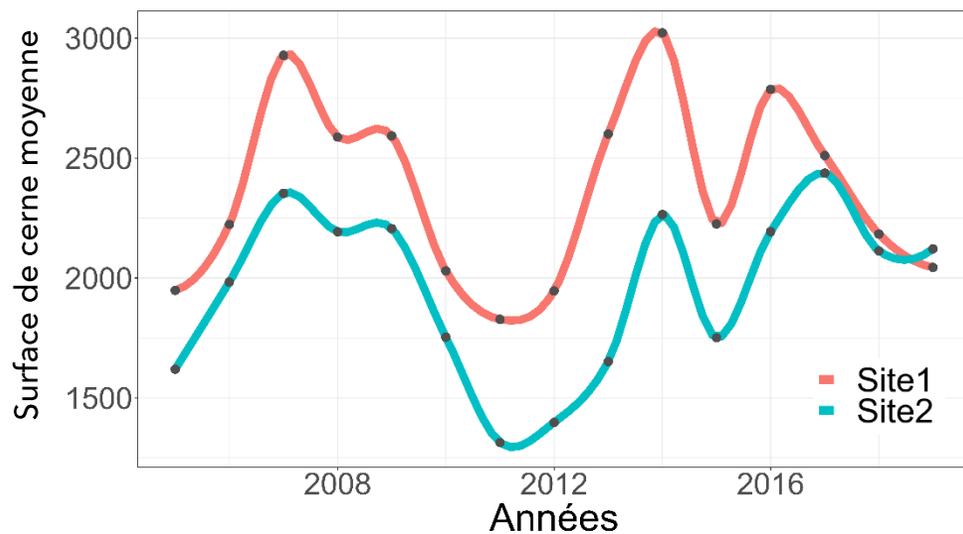
/jour



année

Potentiel de croissance annuel
(2005-2019)

Croissance annuelle et indices environnementaux



Modèle de régression aléatoire

$$y_{ijs} = \sum_{r=0}^{k_m} \varphi_{ijk} m_{sk} + \sum_{r=0}^{k_\alpha} \varphi_{ijk} \alpha_{ik} + \sum_{r=0}^{k_p} \varphi_{ijk} p_{ik} + \varepsilon_{ijs}$$

Trajectoire
phénotypique
individuelle

Trajectoire fixe
moyenne par site

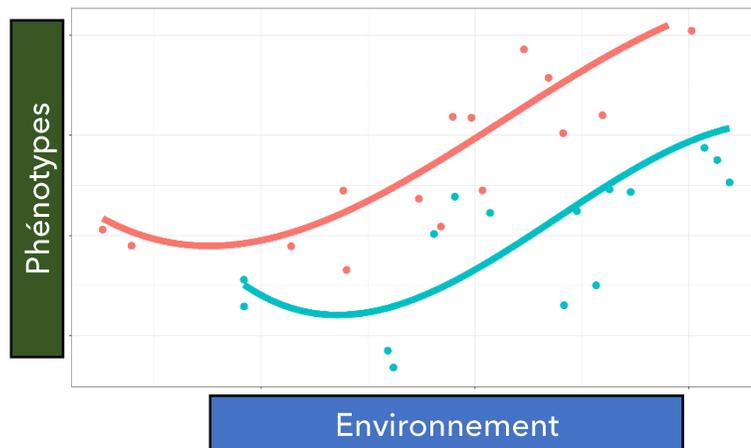
Effet aléatoire
génétique additif

Effet aléatoire permanent
de l'environnement

Avec $\alpha \sim N(0, G \otimes \Omega)$

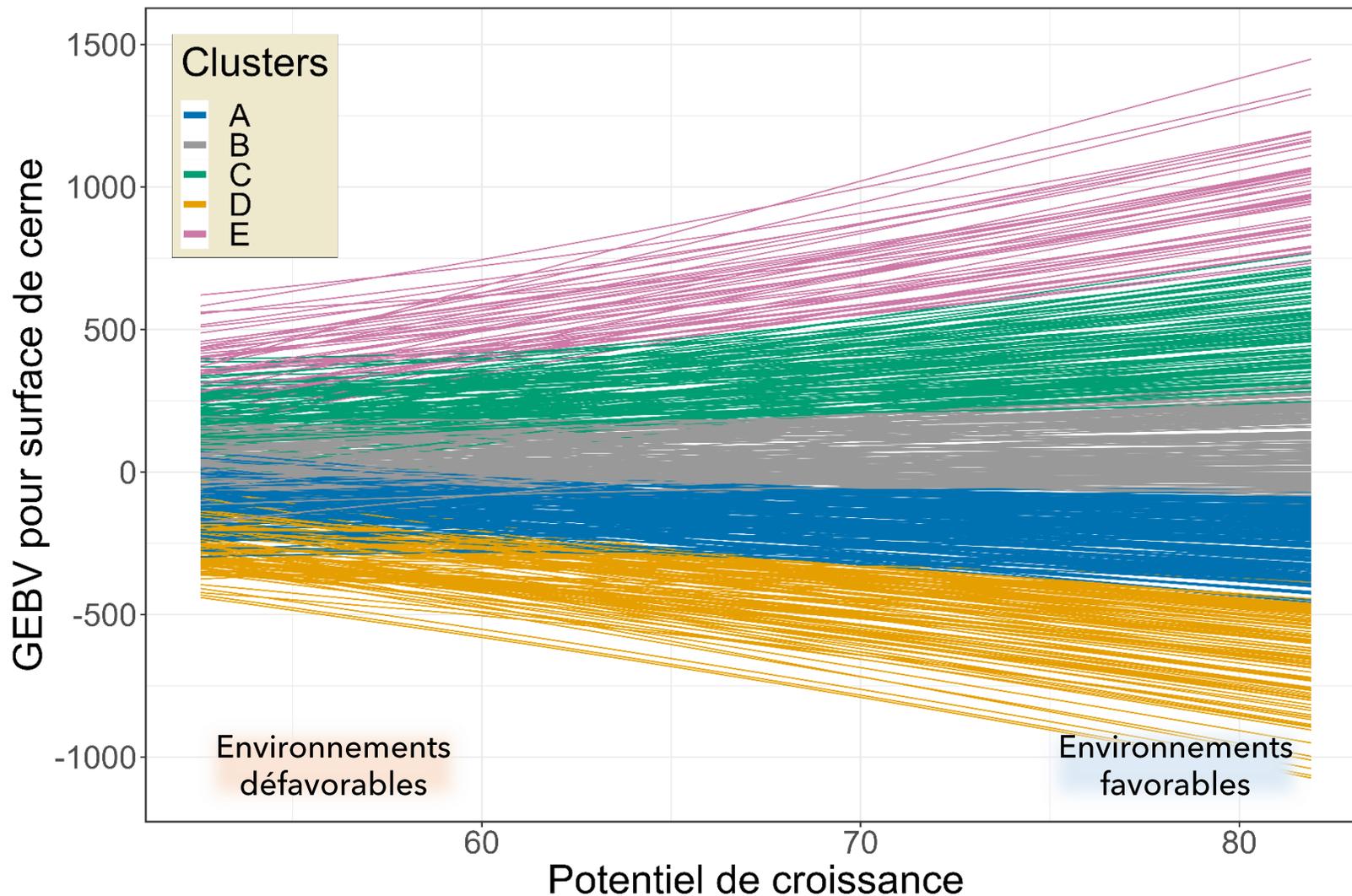
Avec $\gamma \sim N(0, Id \otimes P)$

Génotypage



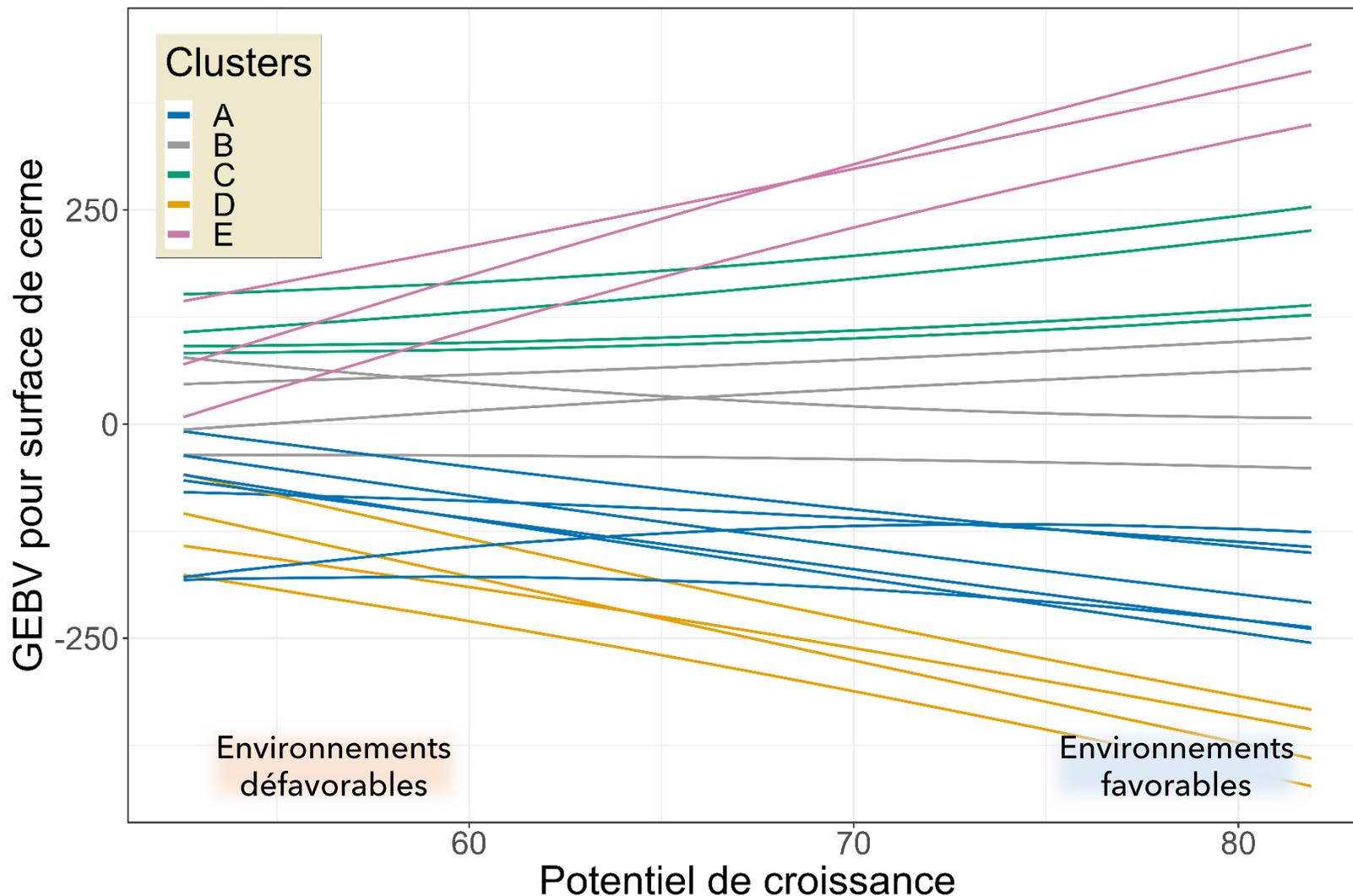
φ_{ijk} : Coefficients polynômes de Legendre

Normes de réaction génomiques



- Représentation de la valeur génétique en fonction du gradient environnemental
- Fortes interactions d'échelle
- Reclassement des génotypes dans les environnements les plus défavorables

Normes de réaction génomiques



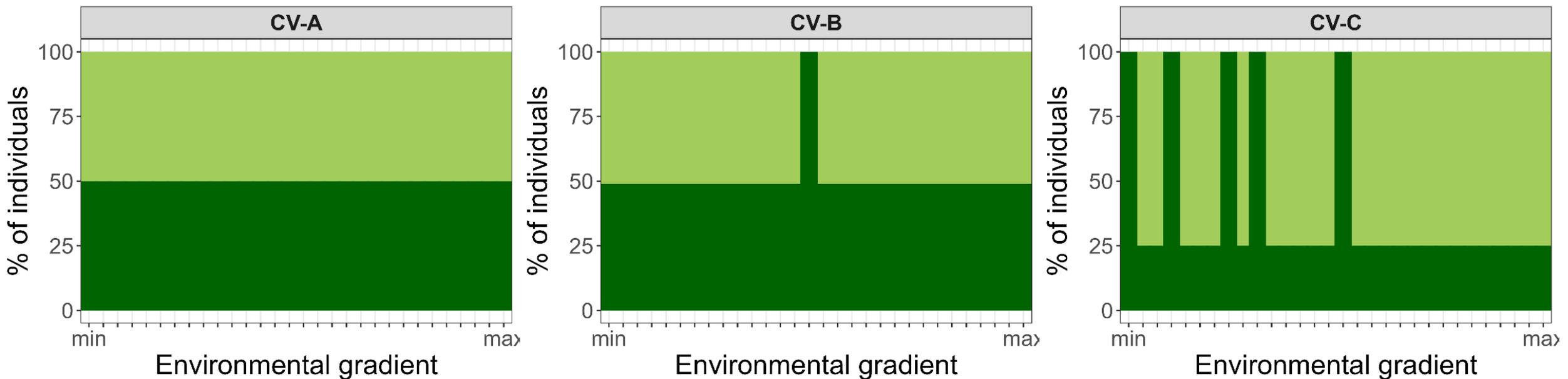
- Représentation de la valeur génétique en fonction du gradient environnemental
- Fortes interactions d'échelle
- Reclassement des génotypes dans les environnements les plus défavorables

Prédictions génomiques en régression aléatoire

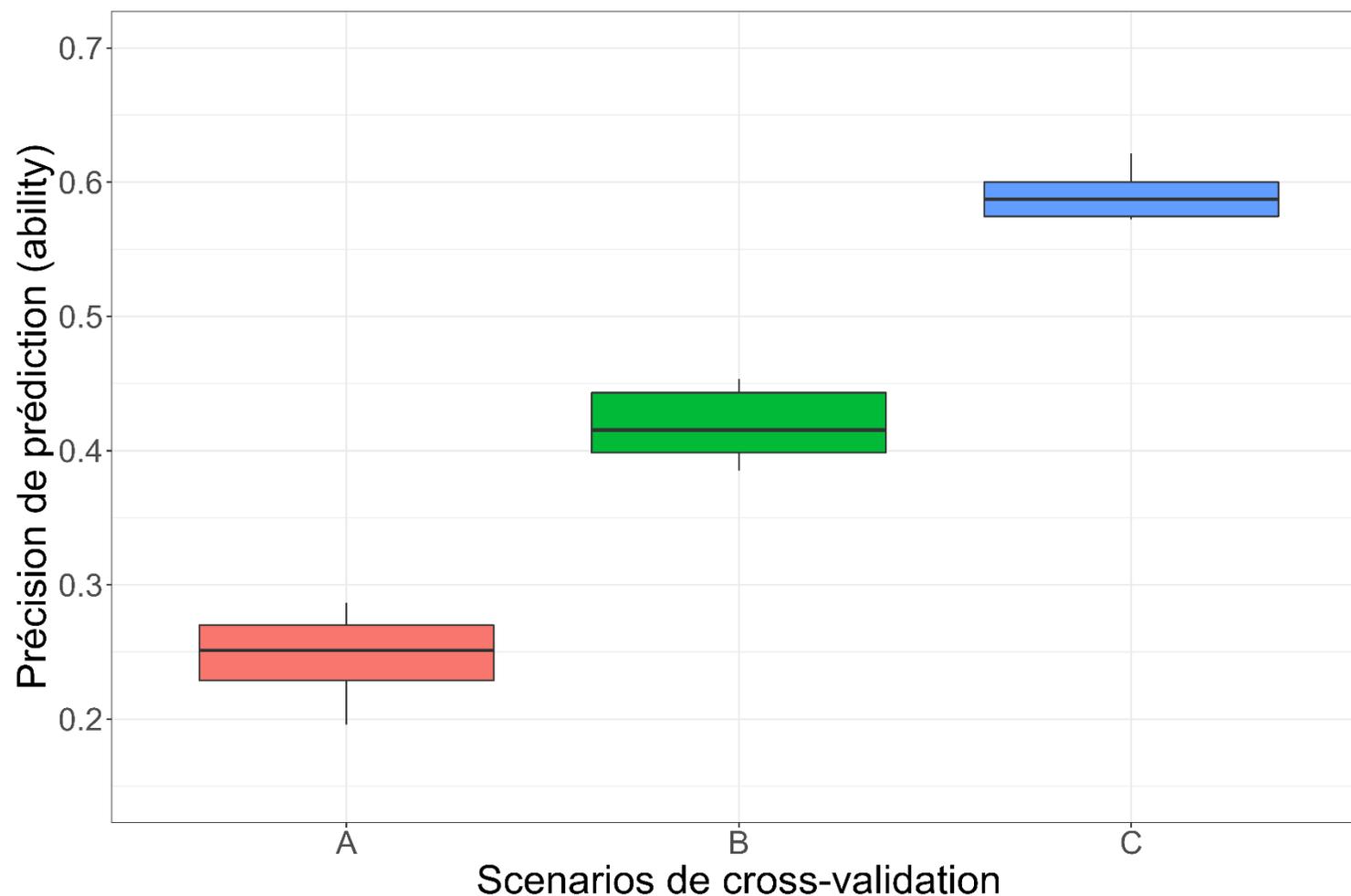
Trois scénarios de cross-validation :

- Même quantité d'information phénotypique en calibration
- Des distributions différentes

■ Pop.validation
■ Pop.calibration



Prédictions génomiques en régression aléatoire



- La distribution de l'information phénotypique est un facteur clé de la précision de prédiction

Conclusion partie n°2

Intégration de l'information environnementale en sélection

- Modélisation de normes de réaction grâce à un phénotypage plus fin
- Choix clé de la variable environnementale
- Re-classement des individus dans les environnements les plus probables à l'avenir

Normes de réaction et sélection génomique

- Distribution de l'information phénotypique clé pour la précision de prédiction
- Intégration de nouveaux traits facilitée par la GS

Conclusion

A photograph of a dense forest of tall, thin trees, likely pines or cypresses, with a path leading through them. The trees are dark and the ground is covered in green ferns. The word "Conclusion" is overlaid in white text on the left side of the image.

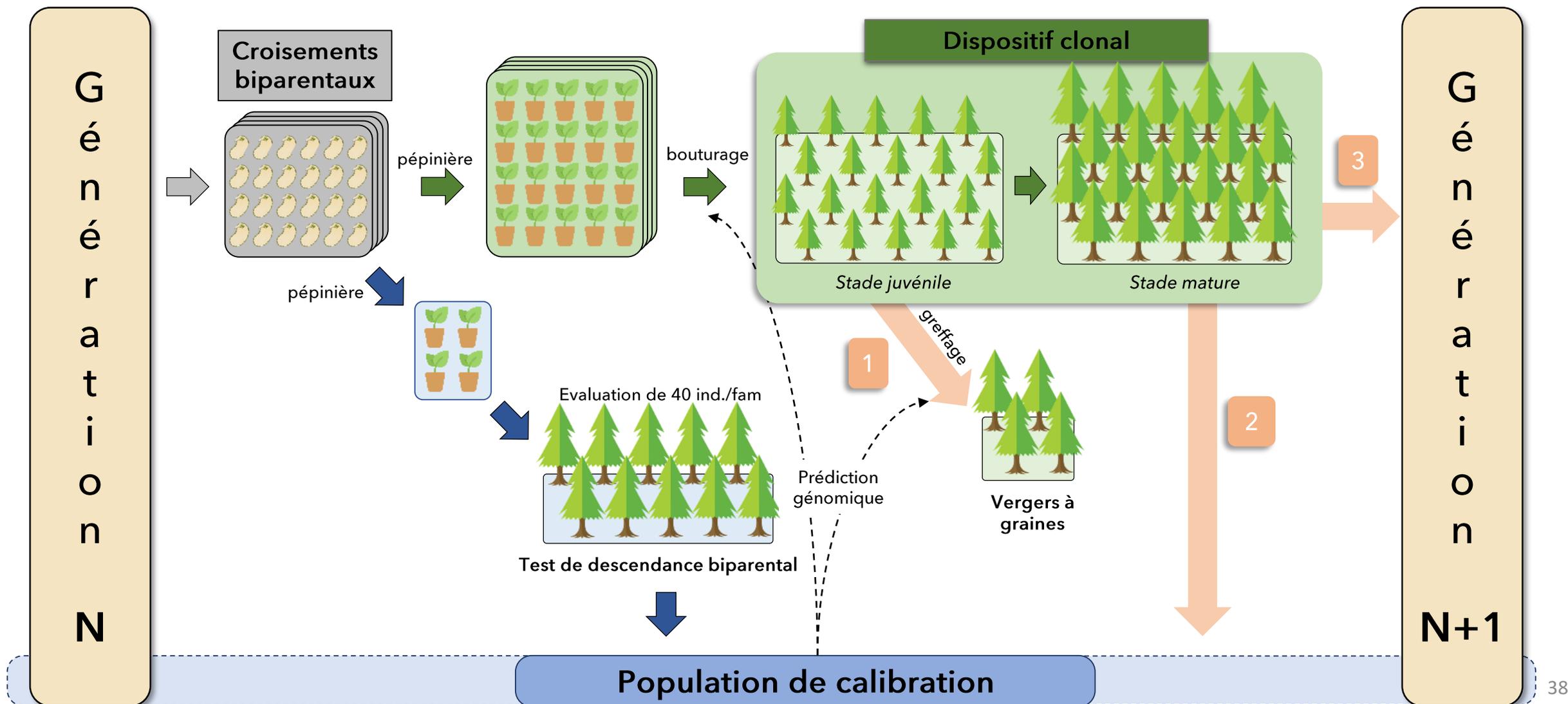
Messages principaux (1)

Sélection génomique possible chez les arbres forestiers

- Précision du modèle génomique devient supérieure à celle du modèle basée sur le pedigree en augmentant la taille des populations de calibration
- L'avantage du modèle génomique coïncide avec une précision intrafamiliale positive
- Maximiser la variabilité génétique des croisements



Schéma intégrant la prédiction génomique



Messages principaux (2)

Changement de paradigme en intégrant l'information environnementale en sélection

- Valoriser des outils et données existantes :
 - Considérer les phénotypes à une échelle plus fine
 - Exploiter les contrastes environnementaux entre années
 - Combiner phénotypes, environnement, et génomique dans un modèle de régression aléatoire
- Sélection génomique doit faciliter l'intégration de traits plus fins et plus complexes



Perspectives de recherche

Intégration des données génomiques

- Approche single-step (matrice H)
- Gestion plus explicite de la diversité



Développement d'outils de phénotypage haut-débit

Caractériser finement et régulièrement la croissance des arbres :

Phénotypage par drones, résistographe, en serre...



Intégration des données environnementales

Comprendre le fonctionnement de l'arbre



→ Collaboration entre disciplines



Remerciements

Directeurs de thèse

- Laurent Bouffier
- Leopoldo Sanchez



Collègues :

- Equipe Xylomes
- Alexandre Bosc
- Ludovic Duvaux
- Mathilde Flores
- Frédéric Lagane
- Céline Lalanne
- Christophe Plomion
- Raphaël Segura

Doc / stagiaires :

- Arnaud Chevalier--Mairet
- Domitille Coq--Etchegaray
- Sukanya Denni
- Thomas Francisco
- Geoffrey Haristoy
- Clément Larue
- Olivier Le Bourdellès
- Gregory Soulet
- Alex Stemmelen
- Adélaïde Theraroz

Highlander Lab :

- Gregor Gorjanc
- Ivan Pocrnic



Comité de suivi de thèse :

- Sophie Bouchet
- Oliver Brendel
- Marie Denis
- Santiago C. González-Martínez
- Gwendal Restoux
- Renaud Rincant

Jury de thèse :

- David Cros
- Jean-Christophe Domec
- Hélène Gilbert
- Anne Laperche
- Timothée Flutre

