



**HAL**  
open science

# Les marqueurs moléculaire et leur utilisation en amélioration des plantes et pour la conservation des ressources génétiques

Marie Pégard, Philippe Barre

## ► To cite this version:

Marie Pégard, Philippe Barre. Les marqueurs moléculaire et leur utilisation en amélioration des plantes et pour la conservation des ressources génétiques. Master. France. 2024. hal-04731290

**HAL Id: hal-04731290**

**<https://hal.science/hal-04731290v1>**

Submitted on 10 Oct 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.



Distributed under a Creative Commons CC0 - Public Domain Dedication 4.0 International License



# Les marqueurs moléculaire et leur utilisation en amélioration des plantes et pour la conservation des ressources génétiques

Marie Pégard et Philippe Barre

## ➤ Plan du cours

1. Rappel sur les marqueurs moléculaires
2. Utilisation des marqueurs dans le cadre de l'amélioration des plantes
3. Utilisation des marqueurs pour la gestion des ressources génétique.



INRAE

## ➤ Rappel sur les marqueurs moléculaires

Voir le CM de Maryse Laloi M1

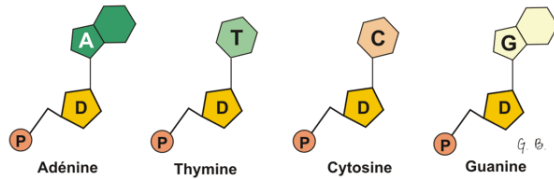
# > L'ADN

## Acide DésoxyriboNucléique

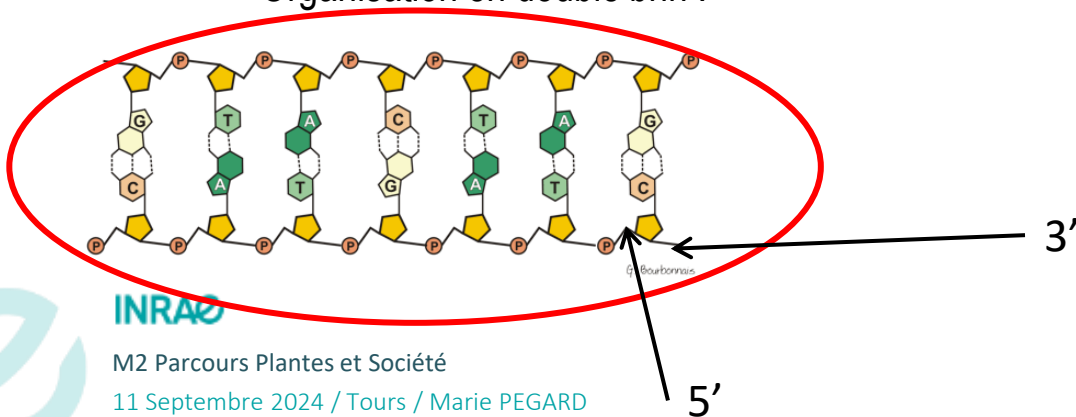
C'est le support de l'information génétique

**Chromosome (métaphase) :** → Forme condensée de l'ADN (= forme visible)

Éléments constitutifs : **4 nucléotides**

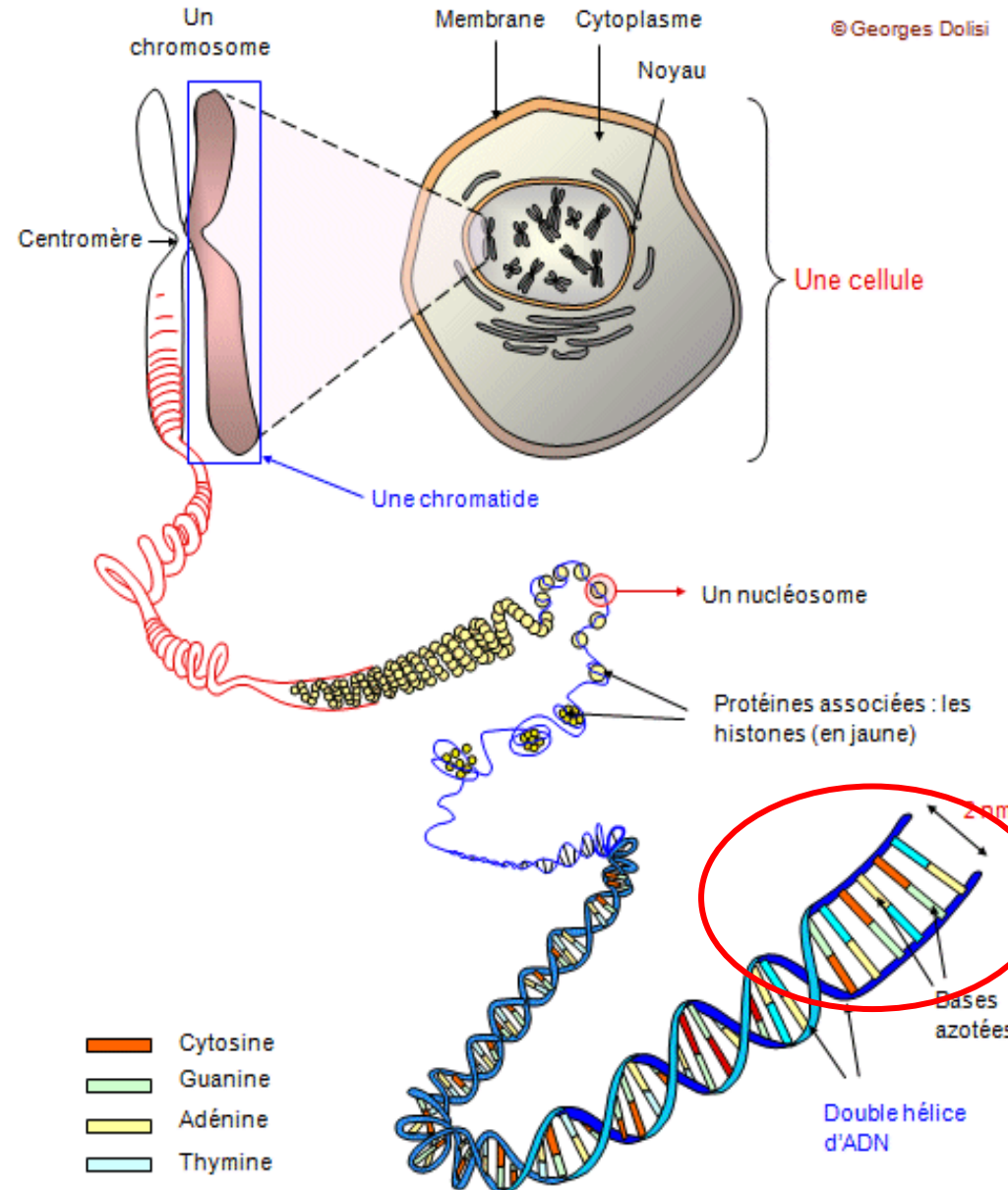
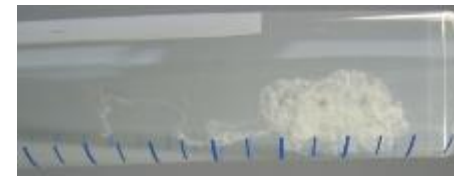


Organisation en double brin :



INRAE

M2 Parcours Plantes et Société  
11 Septembre 2024 / Tours / Marie PEGARD



## ➤ Définition des marqueurs moléculaires

Lié à un **Locus**

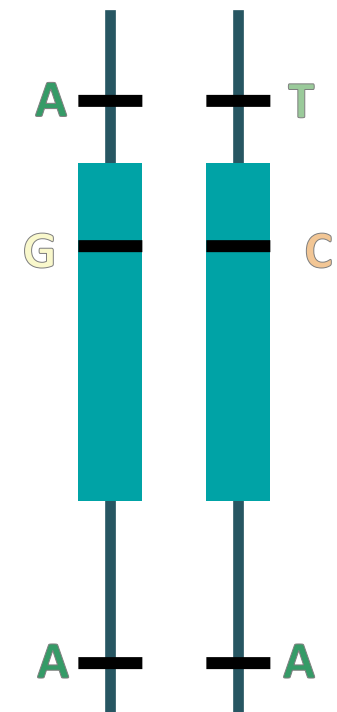
Un endroit sur le génome (fragment d'ADN) : **locus**

De 1 à plusieurs bases

Présentant du polymorphisme entre et au sein des individus : **allèles**

Les différents allèles sont identifiables en totalité (**co-dominant**) ou en partie (**dominant**) pour chaque individu

2 chromosomes homologues



Gène

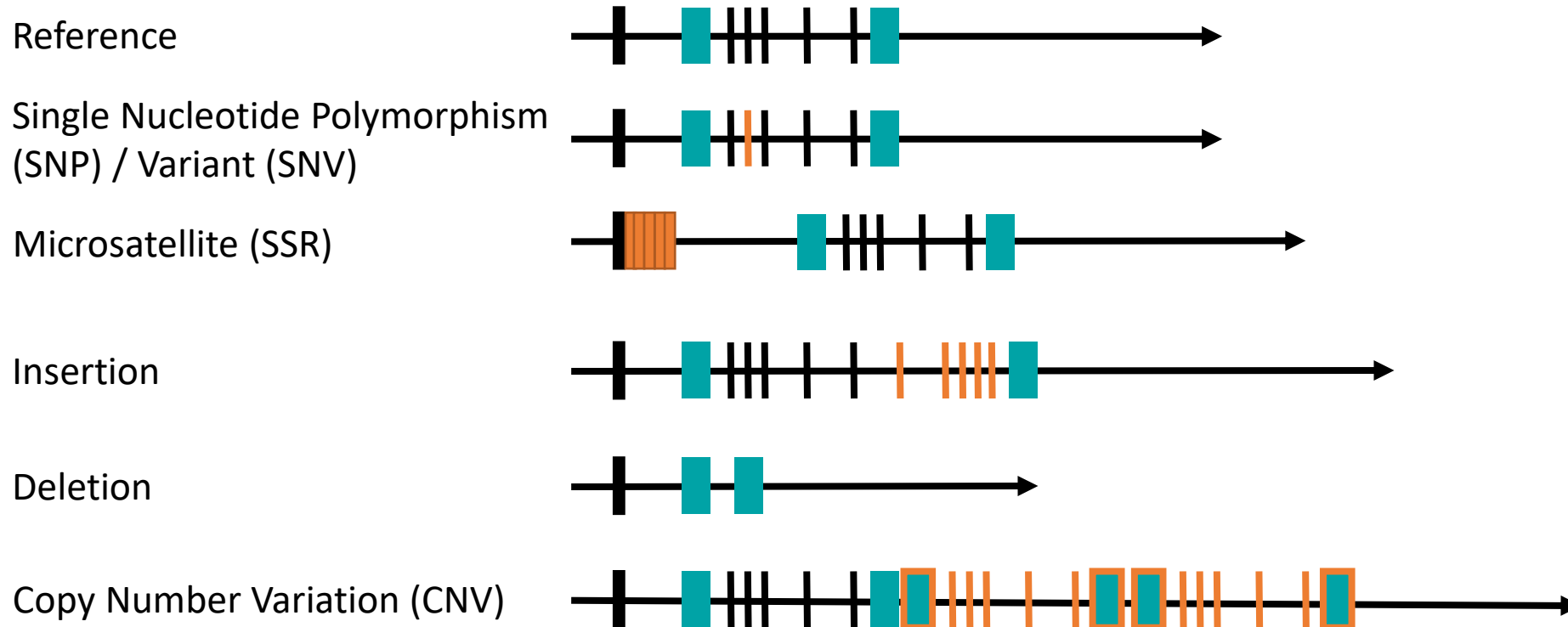
Locus | marqueur

A T G C

Allèles

# ➤ Les marqueurs moléculaires pour quelle mutation?

En fonction de la structure du génome



## ➤ Historique

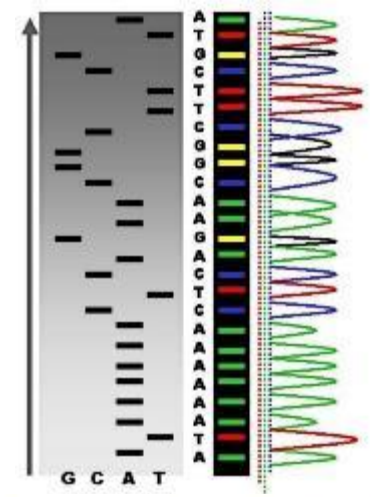
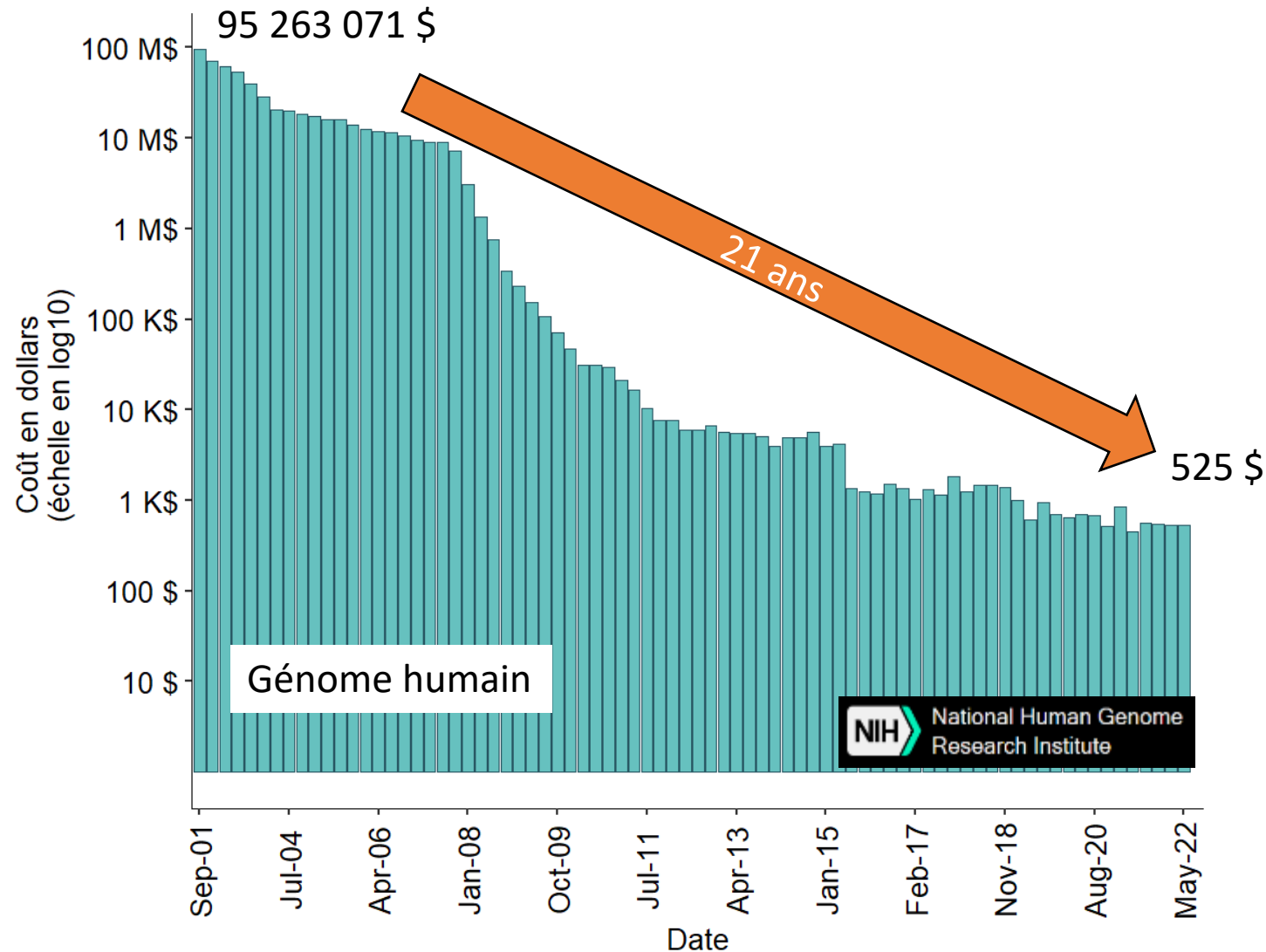
Une évolution fulgurante!!!

- Découverte de la double hélice de l'ADN 1953 (James Watson and Francis Cricks)
- Techniques d'hybridation (années 1970)
- Séquençage de l'ADN 1955 (Fred Sanger)
- Invention de la PCR 1993 (Kary Bank Mullis)



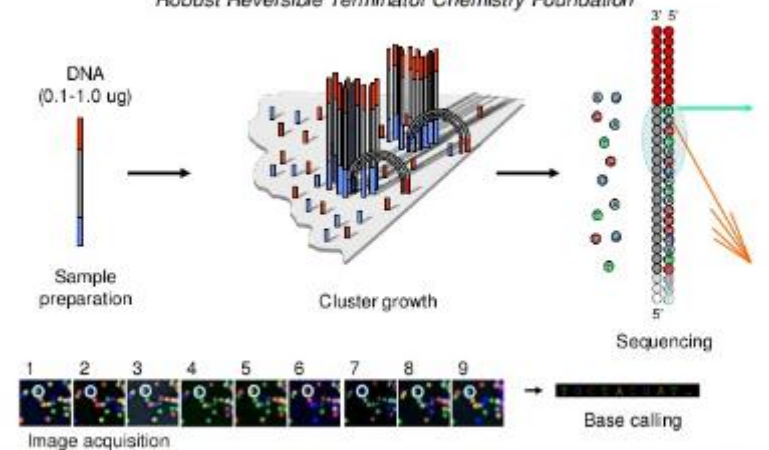
# ➤ Historique

Le séquençage , des coûts de moins en moins élevés



## Illumina Sequencing Technology

Robust Reversible Terminator Chemistry Foundation



- <http://www.ncbi.nlm.nih.gov/genome/probe/doc/Technologies.shtml>
- <http://www2.warwick.ac.uk/fac/sci/lifesci/research/vegin/geneticimprovement/geneticmarker/>
- <http://get.genotoul.fr/>
- <http://gentyane.clermont.inra.fr/>

## ➤ De quoi dépend le coût du génotypage?

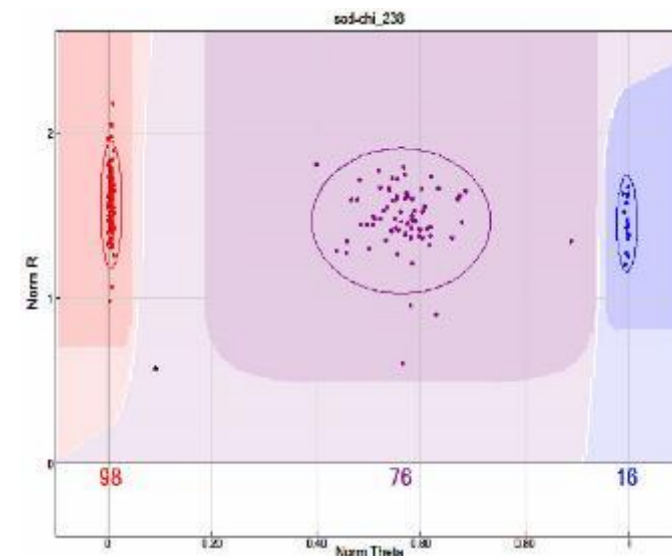
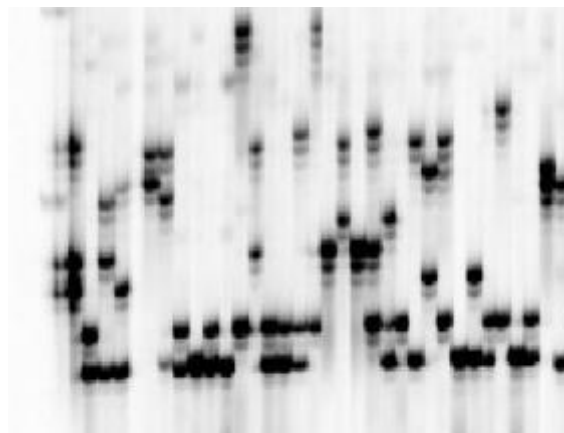
Coût du génotypage dépend de plusieurs paramètres

- De la technologie : réactifs, temps de travail .....
- Du nombre de marqueurs
- Du nombre d'individus

Exemples:

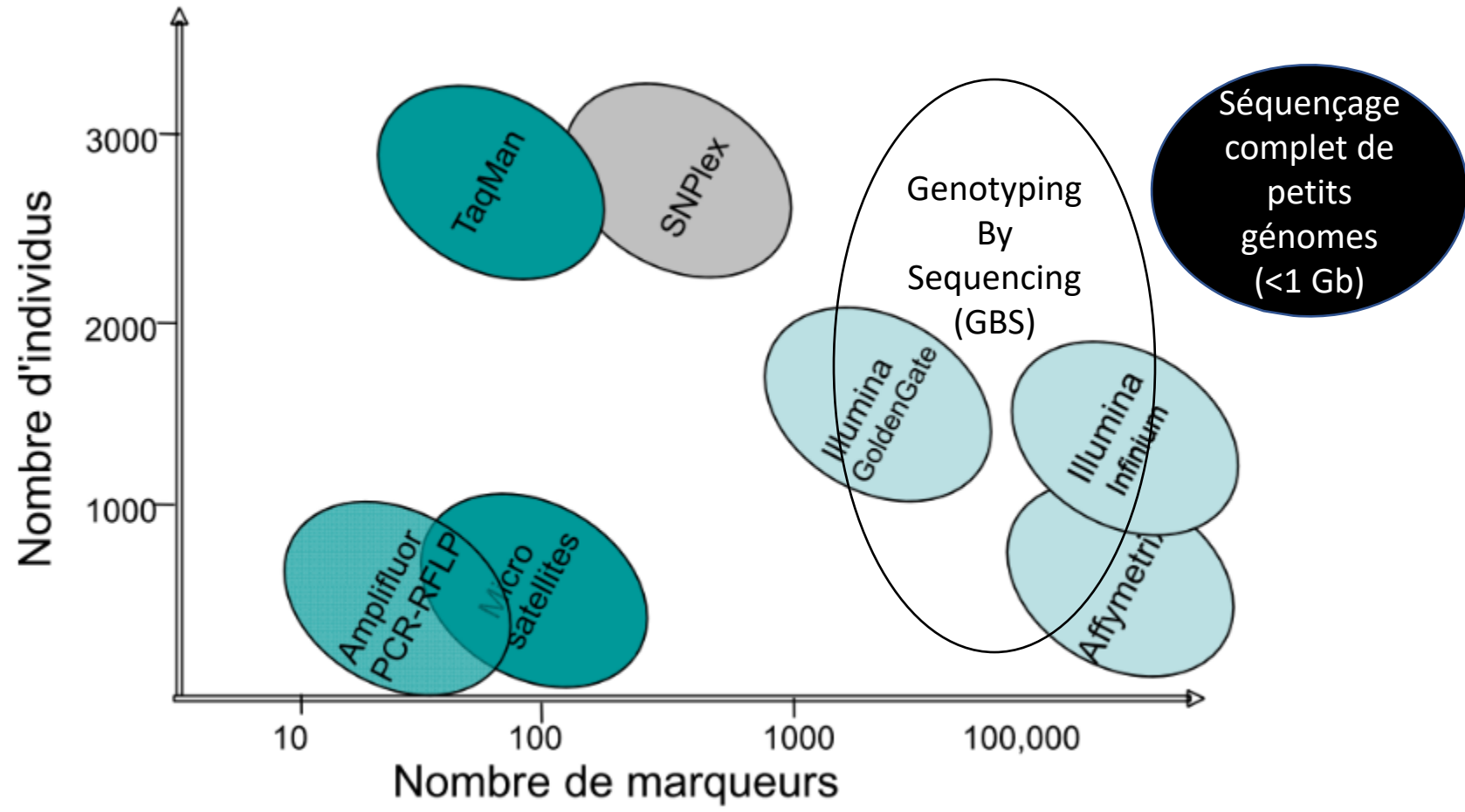
- 0,3 €/point pour un marqueur sur un génotype
- 0,05€/point pour 384 marqueurs sur 1536 individus
- 0,003€/point pour 20000 marqueurs sur 500 individus

Moins le coût du point est cher plus il faut génotyper de marqueurs et d'individus



# ➤ De multiples méthodes de génotypage

Méthode a choisir selon les besoins



## ➤ Génotypage

Quelles régions du génome?

Répartis sur l'ensemble du génome

Ciblés sur des régions particulières:

- Gènes candidats
- Ensemble des gènes

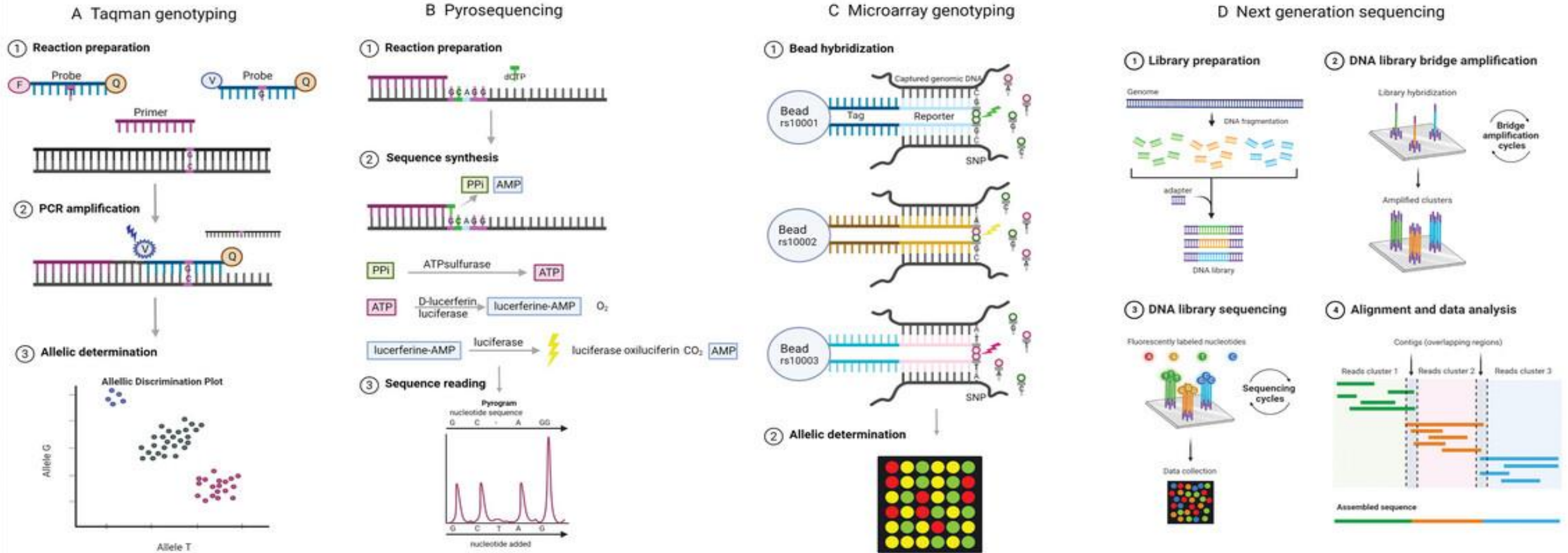
Connaissance préalable du génome

- Séquençage complet / partiel du génome
- Séquençage du transcriptome

Selon les marqueurs à développer, il faut ou non avoir des informations de séquences !!!

# ➤ Rappel des différents types de marqueurs

Pour plus de détail voir le cours M Laloï



INRAE

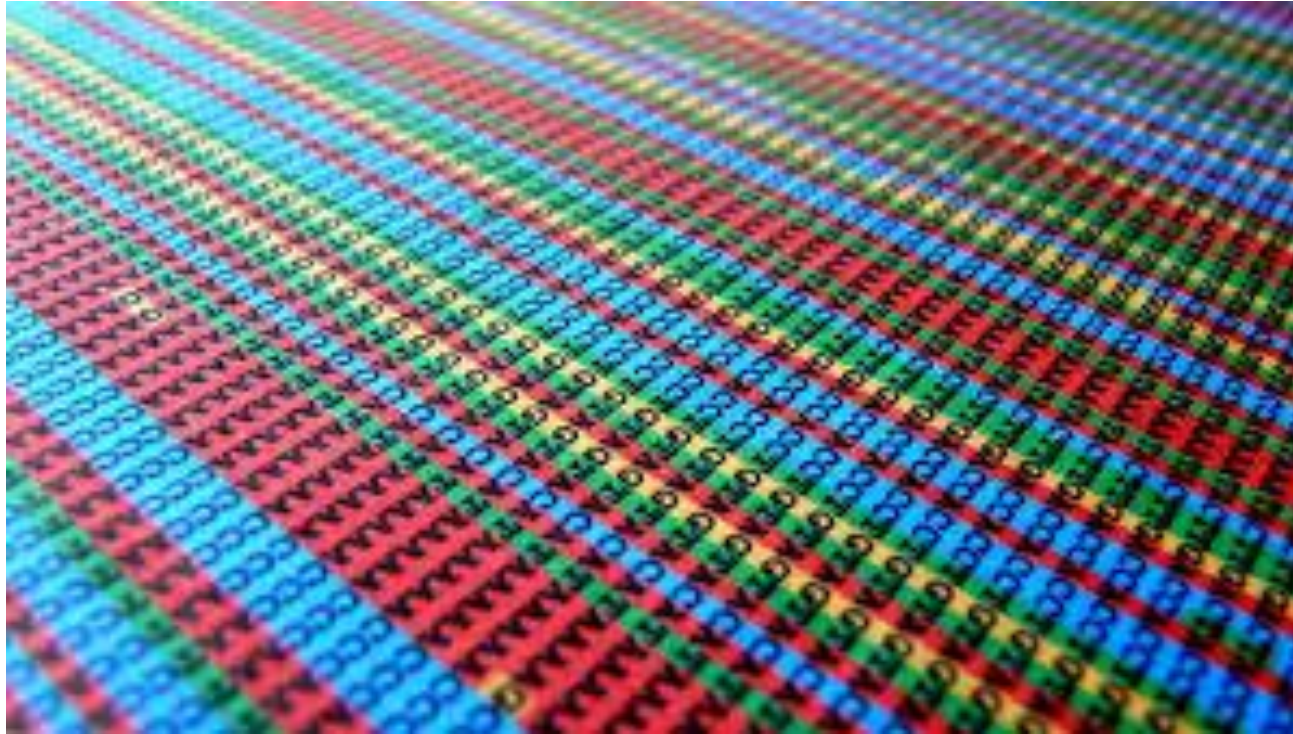
M2 Parcours Plantes et Société

11 Septembre 2024 / Tours / Marie PEGARD

Kockum, I., Huang, J., & Stridh, P. (2023). Overview of genotyping technologies and methods. *Current Protocols*, 3, e727. doi: 10.1002/cpz1.727

## ➤ Résumé des différentes techniques de génotypage

Name	Cost	# markers	Pro	Con	Ref.
PCR-RFLP	+	1	Easy to run in any lab, fast, flexible	Time consuming, manual inspection	(Saiki et al., <a href="#">1985</a> )
Allele-specific PCR	+	1	Easy to run in any lab, fast, flexible	Time consuming, manual inspection	(Gaudet et al., <a href="#">2009</a> )
TaqMan PCR	+	1	Standardized, more accurate	Manual inspection, requires specific equipment	(Hui et al., <a href="#">2008</a> )
Microsatellite	+	1	Robust, do not require specific equipment, flexible	Low resolution, manual inspection, time consuming	(Weber & May, <a href="#">1989</a> )
Pyrosequencing	++	1	Captures all potential alleles	Time consuming, manual inspection, requires specific equipment	(Kreutz et al., <a href="#">2013</a> )
iPLEX	++	40	Multiplex assay	Manual inspection, requires specific equipment	(Gabriel et al., <a href="#">2009</a> ; Tang et al., <a href="#">1999</a> )
Multiplex TaqMan	++	up to 100	Multiplex assay	Requires specific equipment	(Martínez-Cruz et al., <a href="#">2011</a> )
Genotyping arrays	+++	50k-2 mil	Multiplex assay	Requires specific equipment and expertise	(Verlouw et al., <a href="#">2021</a> )
NGS-Exome	++++	25000	captures all coding variants	Requires specific equipment and expertise, demanding processing	(Seaby et al., <a href="#">2016</a> )
NGS-Whole Genome	+++++	up to 40 mil	captures all variants	Requires specific equipment and expertise, demanding processing	(Slatko et al., <a href="#">2018</a> )



## ➤ FOCUS SUR LE GÉNOTYPAGE PAR SÉQUENÇAGE

## ➤ Génotypage par séquençage

Séquencer tout le génome (WGS : Whole genome Sequencing):

➤ oui mais peut être coûteux selon la taille du génome

Espèce	Taille génome	\$ Coût estimé pour 30X
Arabidopsis	120 Mb	20.98
Humain	3000 Mb	525
Blé	16 Gb	2797

Séquencer une partie du génome :

Les gènes : transcriptome

Des régions réparties sur le génome mais non « choisies » : GBS avec enzyme de restriction

Des régions choisies : capture



## ➤ Génomique par séquençage

Trois étapes:

1. Fabrication de la « banque » : mélange de fragments d'ADN : peut être réalisée dans un labo de biologie moléculaire classique + compétence BM
2. Séquençage : nécessité des machines très chères et qui évoluent très vite
3. Analyse bio-informatique : nécessite des ordinateurs très puissants + compétences en bio-informatique



## ➤ Les banques

Banques d'ADNc : attention aux tissus choisis selon la question et si besoin standardisation

Banque d'ADN génomique :

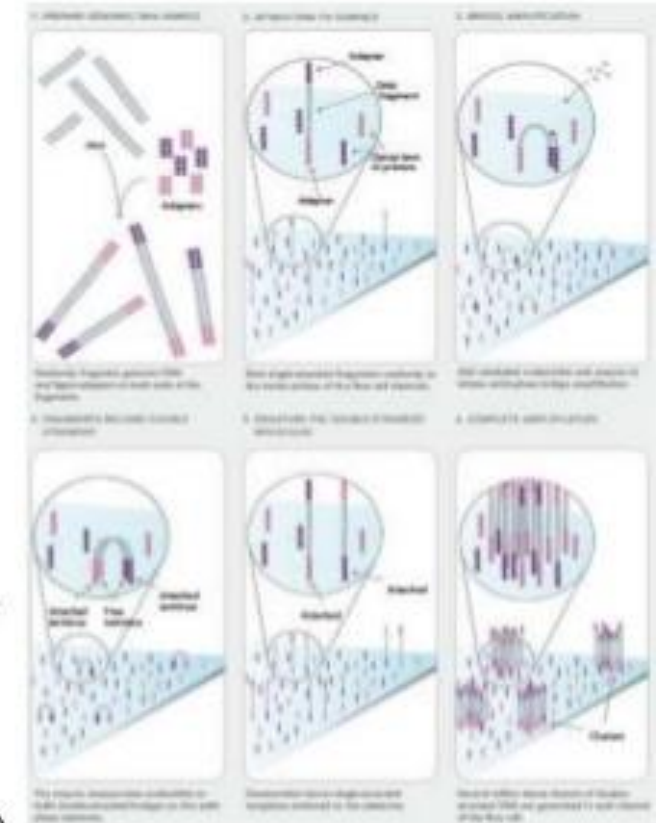
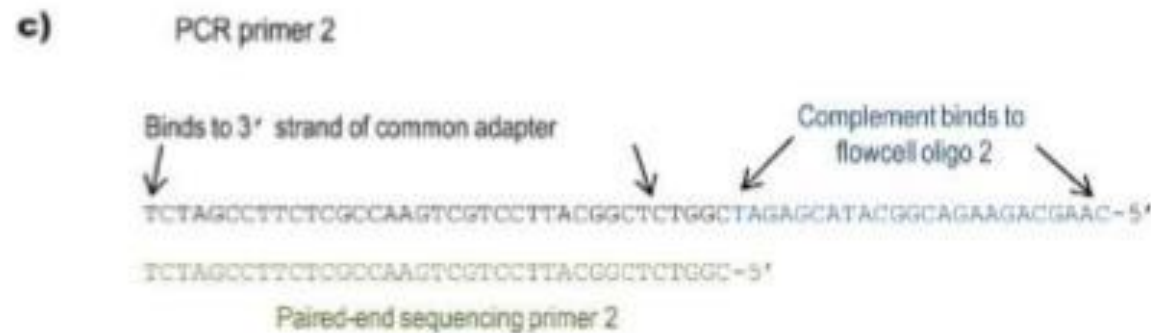
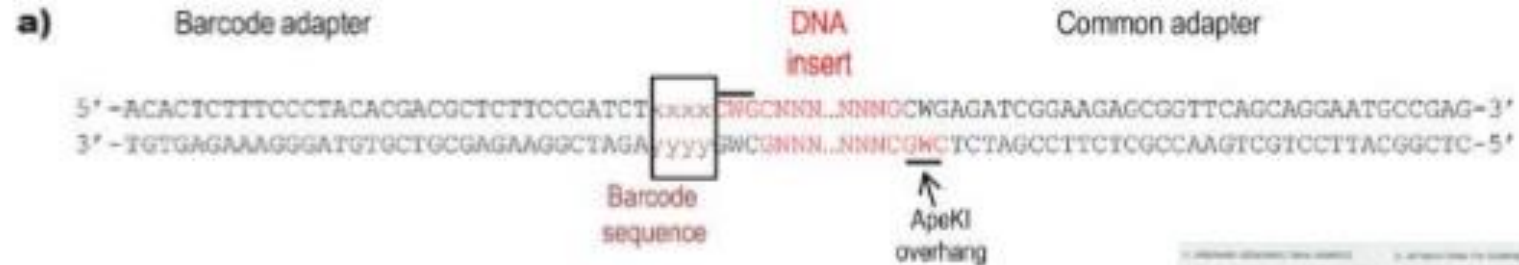
- ADN total fragmenté de manière aléatoire

- ADN issu de l'amplification d'une partie du génome par PCR: séquences cibles (ex. barcoding) ou enzymes de restriction

- ADN issu de l'hybridation sur des séquences connues : capture

# ➤ Exemple : GBS Elshire et al.2011 PLOS One

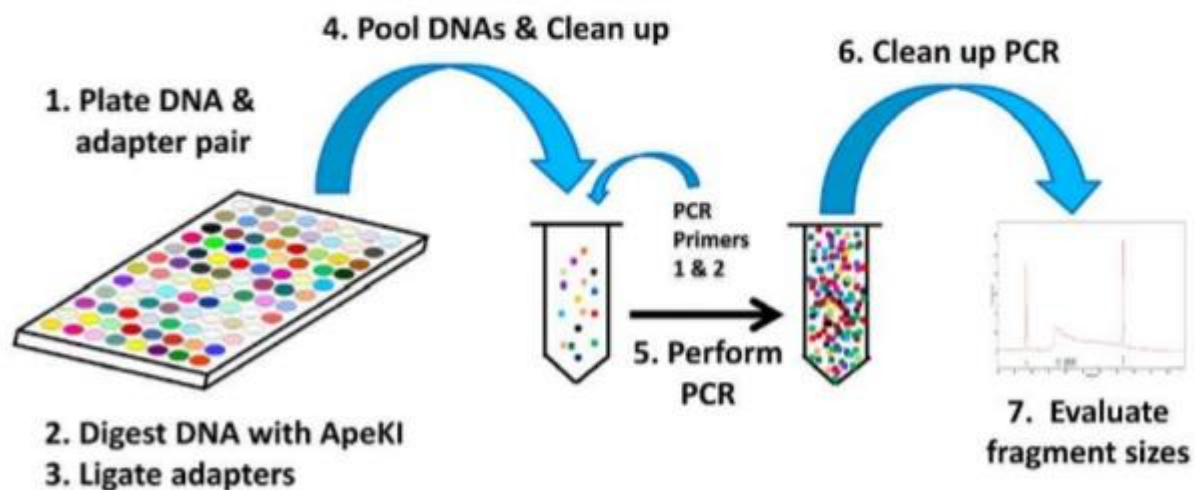
Digestion de l'ADN génomique par une enzyme de restriction (ici ApeKI)



(Elshire et al., 2011 PLOS One)

## ➤ Exemple : GBS Elshire et al.2011 PLOS One

### GBS library construction



Il existe beaucoup de variantes RADseq....

(Elshire et al., 2011 PLOS One)

# ➤ Séquençage: plusieurs plateformes selon les besoins

Séquenceurs 2 <sup>ème</sup> génération											
Société	Roche			Illumina				Life Technologies			
Plateforme											
Technologie	Titanium	FLX Titanium FLX +						Chip 314 Chip 316 Chip 318			
Acides nucléiques (matrice)											
Ligation adaptateurs											
Méthode d'amplification	 PCR en émulsion			 « Bridge PCR »				 PCR en émulsion			
Méthode de séquençage	Synthèse (Pyroséquençage)			Synthèse				Ligation			
Durée de séquençage/run	10h	10h 20h		26h	8jrs	8jrs	14jrs	2h	12jrs	8jrs	8jrs
Capacité (Mb) séquençage/run	50	500 900		1500	100000	200000	95000	>10 >100 >1000	70000	80000	150000
Taille moyenne des reads	400	400 700		150+150	100+100	100+100	150+150	100 >100 >100	50+35	75+35	75+35
Coût (\$) /run	1100	6200		750	10000	20000	11500	500 750 950	8150	6100	10500
Coût machine + annexes ((K\$))	110+25	500+30		125	560	690	250	50+20	480+55	350+55	600+55
Exactitude de séquençage (%)	99	99		99,9	99,9	99,9	99,9	99	99,95	99,95	99,99

Nouveaux séquenceurs encore plus performants



## ➤ Bio-informatique

Enormément de données à traiter : penser à la mémoire !!! : machines dédiées

Logiciels :

Quelques uns sous Windows dont CLC genomics: convivial, payant, peu modulable

Beaucoup d'application sous Linux : gratuit, très modulable, moyennement convivial

Un système intermédiaire : Galaxy <https://usegalaxy.org/>

## ➤ Bio-informatique : plusieurs étapes

Démultiplexage : séparation des séquences par génotypes

Nettoyage : enlever les adaptateurs, bonne qualité, longueur minimale

Alignement des séquences:

- Sur une séquence de référence

- Sur un assemblage de novo

Nettoyage post-alignement

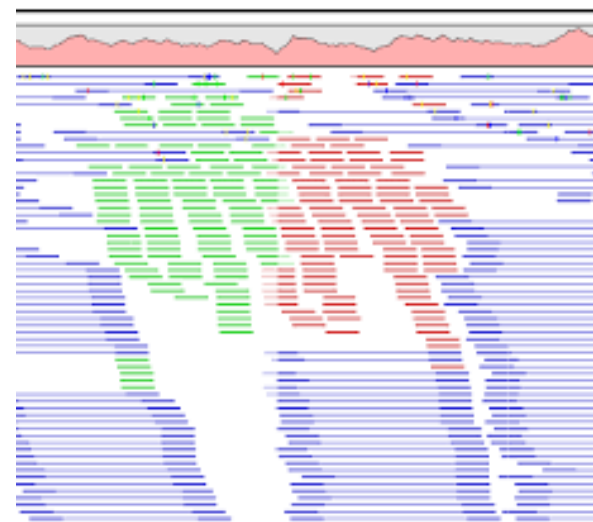
- Duplicats

- Non aligné

Détection des SNP

- Profondeur de séquence

- Fréquence minimale d'un allèle



# ➤ Optimisation du coût de génotypage

## L'imputation génotypique

Décrite pour la première fois par Burdick et al., 2006 sous le terme « in silico genotyping »

Renvoi au processus de prédire la valeur du génotypage non disponible au locus pour un individu

Deux méthodologies reposant sur :

La relation entre les individus via l'apparentement et la ségrégation mendélienne

Peuvent être combinées

SNPs

1	0	1	2	0	0	1	2	0	1	0
2	1	2	0	0	1	1	0	2	0	2
				...						
43	0	2	2	?	?	1	0	?	?	1
44	1	2	0	?	?	1	2	?	?	0
				...						
1039	2	0	1	?	?	0	1	?	?	2

La relation entre les marqueurs : le déséquilibre de liaison (LD)



# ➤ L'imputation génotypique

Le principe

Chaque échantillon phasé et les haplotypes sont modélisés comme une mosaïque

Données génotypiques manquantes sur certains SNP (points d'interrogation gris)

1	?	?	?	1	?	1	?	0	2	2	?	?	2	?	0
0	?	?	?	2	?	2	?	0	2	2	?	?	2	?	0
1	?	?	?	2	?	2	?	0	2	1	?	?	2	?	0
1	?	?	?	2	?	1	?	1	2	2	?	?	2	?	0
2	?	?	?	2	?	2	?	1	2	1	?	?	2	?	0
1	?	?	?	1	?	1	?	1	2	2	?	?	2	?	0
1	?	?	?	2	?	2	?	0	2	1	?	?	2	?	1
2	?	?	?	1	?	1	?	1	2	1	?	?	2	?	1
1	?	?	?	0	?	0	?	2	2	2	?	?	2	?	0



1	1	1	1	1	2	1	0	0	2	2	0	2	2	2	0
0	0	1	0	2	2	2	0	0	2	2	2	2	2	2	0
1	1	1	1	2	2	2	0	0	2	1	1	2	2	2	0
1	1	2	0	2	2	1	0	1	2	2	1	2	2	2	0
2	2	2	2	2	1	2	0	1	2	1	1	2	2	2	0
1	1	1	0	1	2	1	0	1	2	2	1	2	2	2	0
1	1	2	1	2	1	2	0	0	2	1	1	1	2	1	1
2	2	2	1	1	1	1	0	1	2	1	0	1	2	1	1
1	2	2	0	0	2	0	0	2	2	2	1	2	2	2	0

Comparaison avec des haplotypes de référence (provenant d'un génotypage plus dense)

0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0
0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

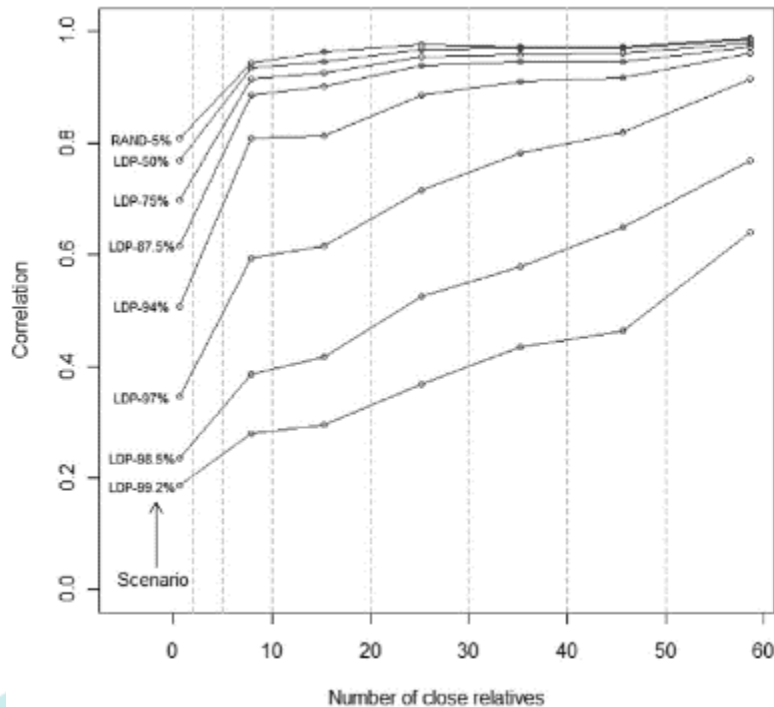
Imputation (prédiction) des allèles dans les échantillons pour créer des génotypes imputés (orange)



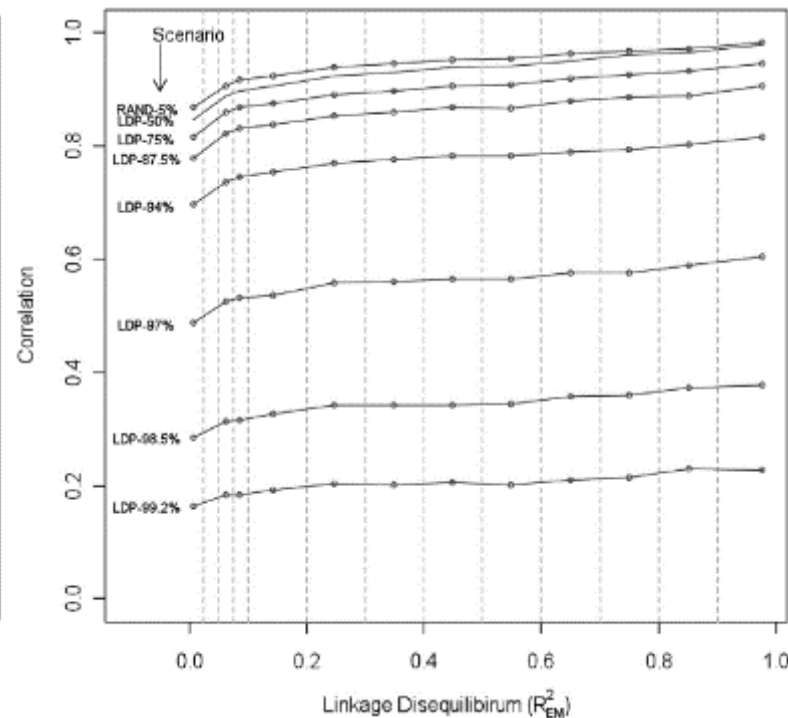
# ➤ L'imputation génotypique

Quels facteurs influence la qualité de l'imputation

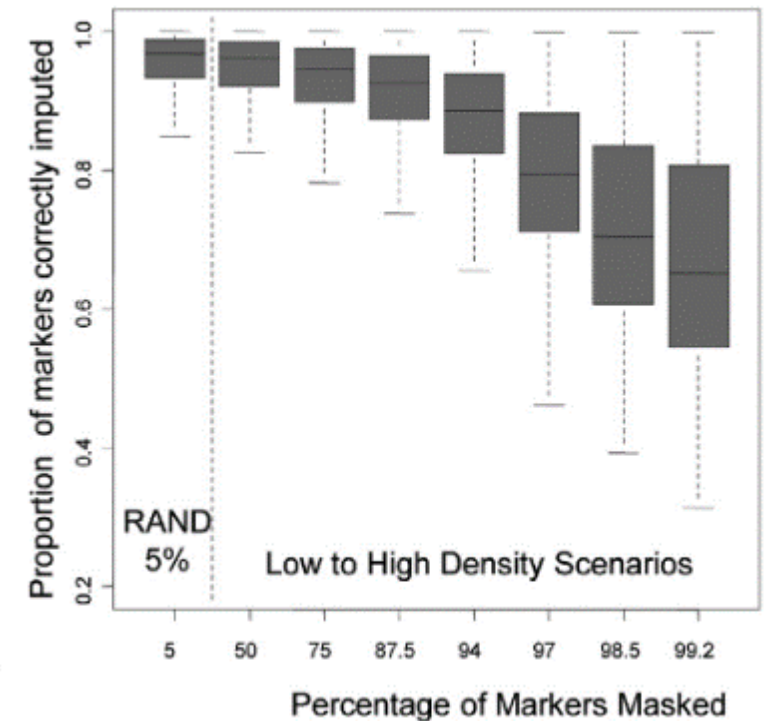
L'apparentement entre les populations de référence et les populations imputées



le niveau de déséquilibre de liaison (DL)



la densité des marqueurs qui influe sur la perception de la liaison.



# ➤ Optimisation du coût de génotypage

Exemple avec l'imputation

Peuplier noir : 1039 individus

- 35 familles venant de 23 parents
- Tous génotypés avec une puce 12K (7540 SNPs)
- 43 individus séquencés :
  - 23 parents
  - 17 descendants
  - 6 non apparentés

Table 1: Number of individuals and pedigree information

	Male	Factorial mating				Multiple Pair mating								Total genotyped	Total Sequenced
		SAN-GIORGIO	SRZ	BDG	71077-308	92510-1	72145-007	72131-017	73182-009	73193-056	72131-036	3824-3	71034-2-406		
Female														12	12
Factorial mating	VGN-CZB25	55	57 [4]	54 [1]	32		31	14	30					277	6
	<u>71041-3-402</u>		28	11 [1]	17	30		25						112	2
	71072-501	25	28	29 [1]										83	2
	SSC	15	20	20	22 [1]									78	2
Multiple Pair mating	71040					24 [1]	20							45	2
	<u>66220037</u>					25	32 [1]	118	31					207	2
	73193-089						22	20	18 [1]					61	2
	<u>662200216</u>							31 [1]		19				51	2
	71069-914										22 [1]			23	2
	73193-091										21		30 [1]	52	2
II480										13	19 [1]		33	2	
Population	Unrelated													6	6
Total genotyped	17	97	134	115	72	80	109	209	80	20	57	20	31	1039	
Total Sequenced	17	1	5	4	2	2	2	2	2	1	2	2	2		43

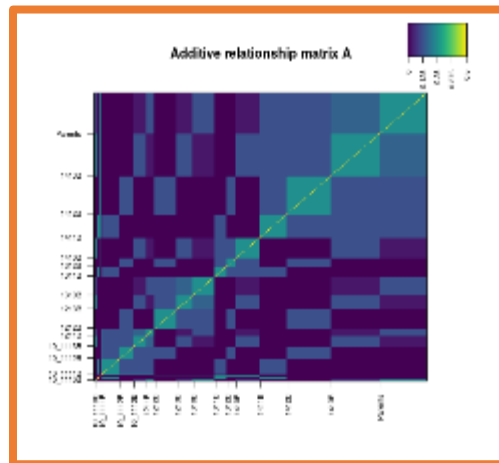
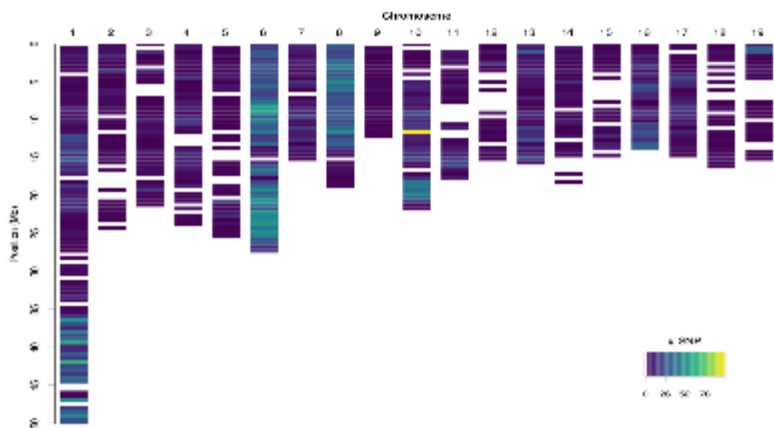
Description of black poplar breeding resources used in the study, with mating designs involved and number of individuals per family in the inner cells. Parental and family cells are coloured by class in the mating regimes: yellow, factorial mating progenies; orange, multiple pair mating progenies; red, factorial mating parents; purple, multiple pair mating parents; and dark cyan, unrelated individuals. In brackets, some selected cells show the number of sequenced progenies, with the figure in red involving 2 progenies that were subsequently used as parental females (underlined codes) for the multiple pair mating.

# ➤ Optimisation du coût de génotypage

Exemple avec l'imputation

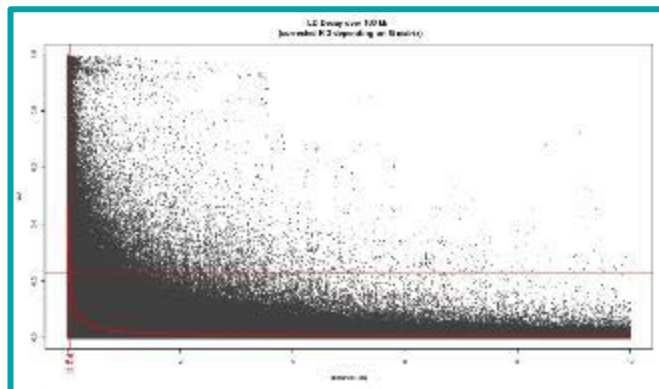
Basse densité : Puce 12K SNP Infinium (Faivre-Rampant et al., 2016)

- Répartition non homogène
- Plus de marqueurs dans les zones avec des QTL connus

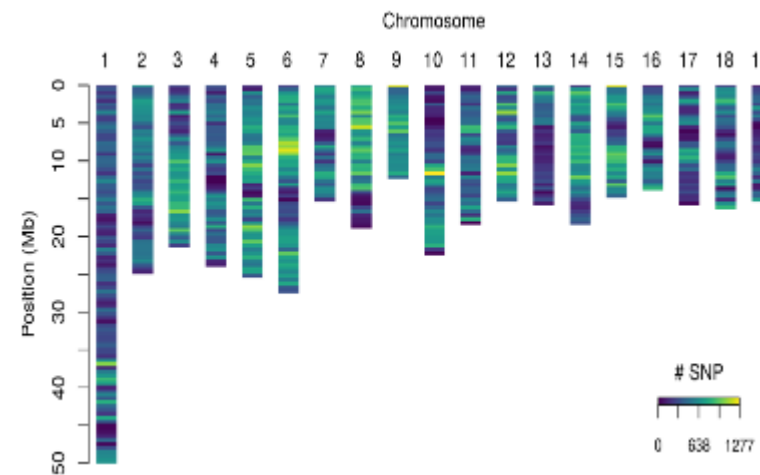


Imputation avec Fimpute<sup>2</sup>

- **Pédigrée information**
- **Population information**



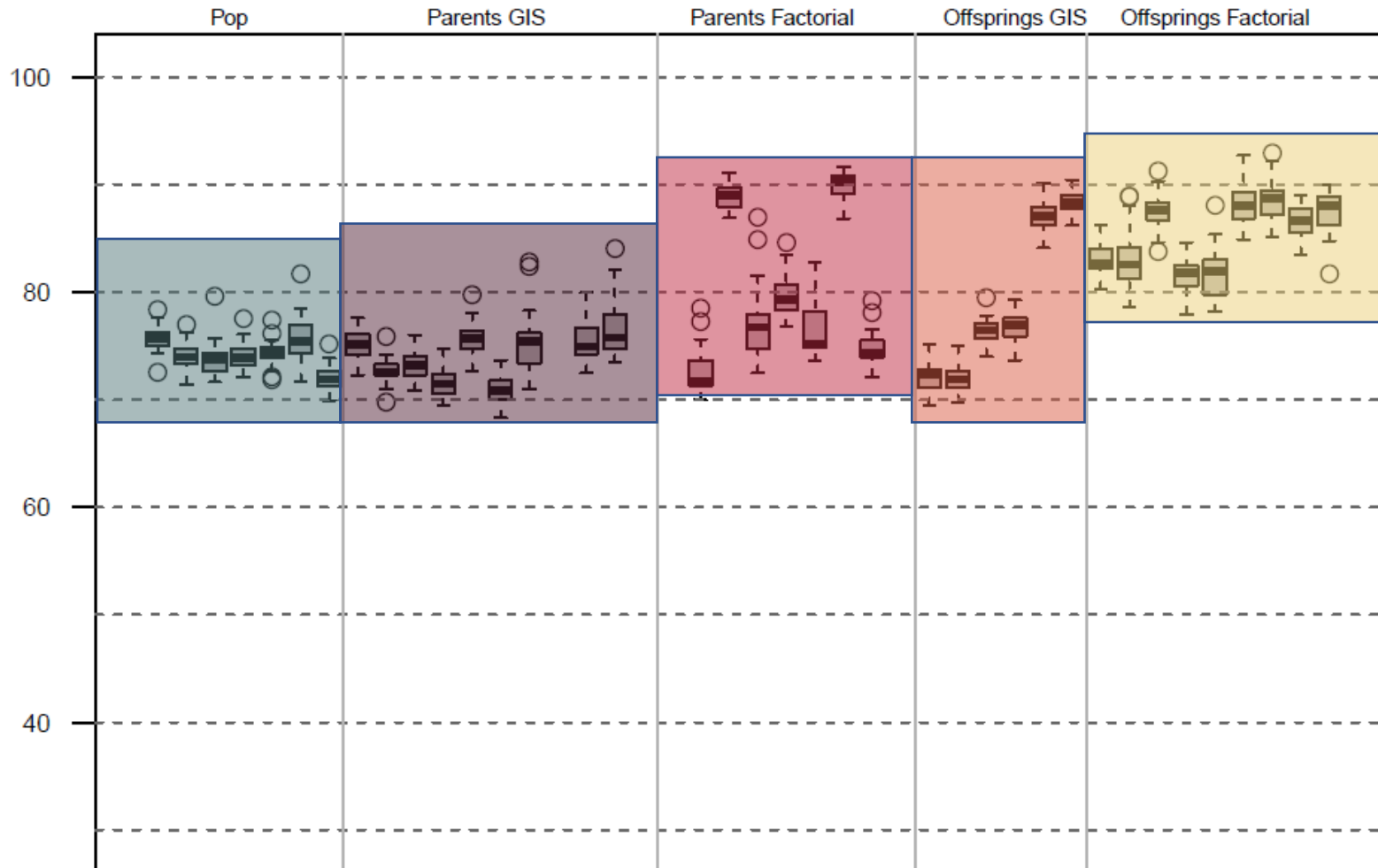
Haute densité : WGS => 1.4 Millions SNP



# ➤ Optimisation du coût de génotypage

Exemple avec l'imputation

**Percentage of correctly imputed SNP by individuals**



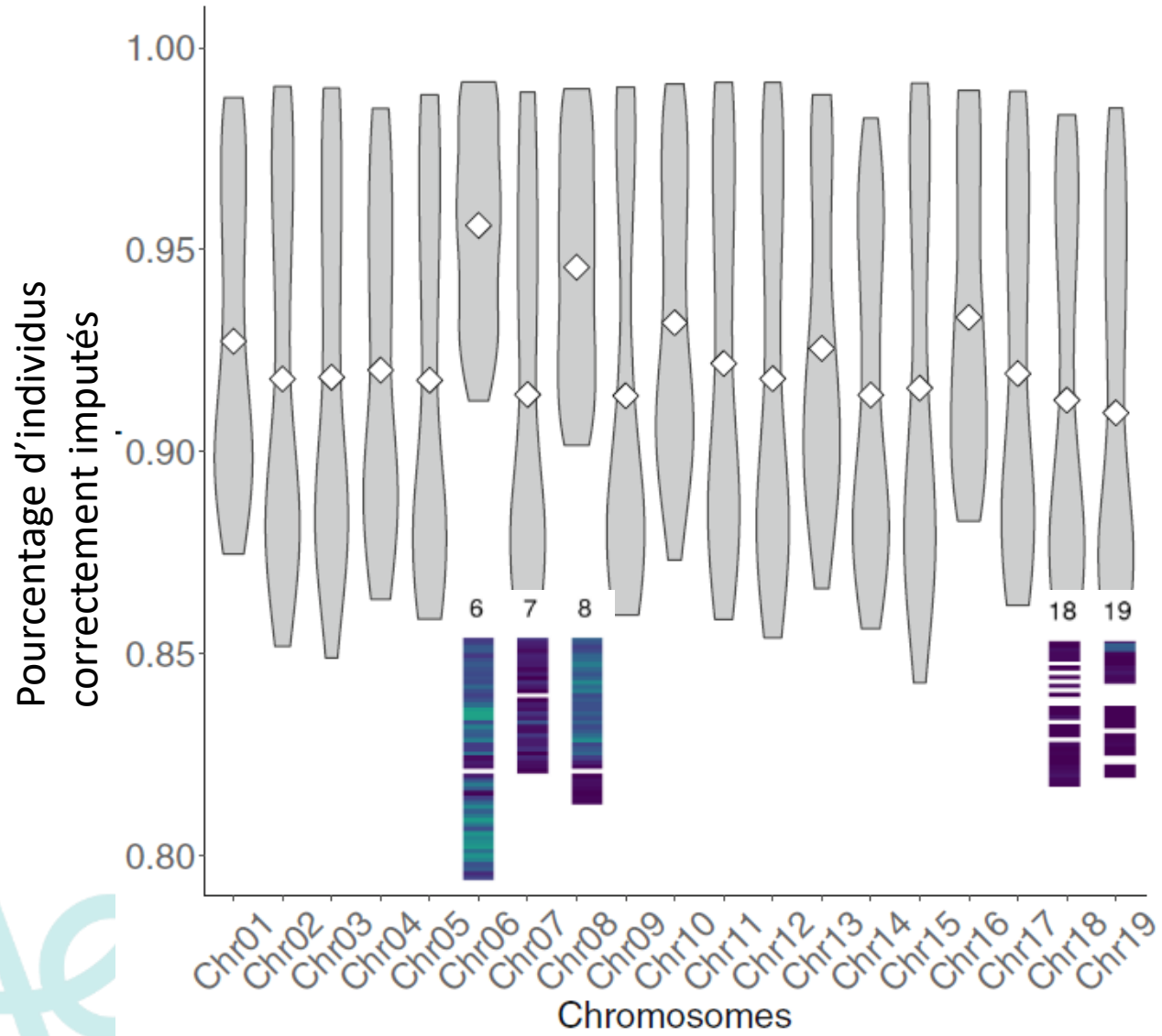
Impact de l'apparentement et du nombre d'apparentés sur la qualité de la prédiction

Table 1. Number of individuals and pedigree information

	Male	Pedigree mating											
		Factorial mating						Multiple Pair mating					
Female		SRZ	EDC	7077-308	6950-1	7245-07	7211-017	7302-00	7101-066	7211-036	182-3	7104-1-06	7216-1
Factorial mating	7074-0225	22	27	31	32		31	11	11				
	7211-0-016		28	11	17	21		26					
	7102-501	28	28	29									
	850	18	21	21	22								
Multiple Pair mating	71010					25	21						
	6822017					25	22	18	11				
	7312-149						22	20	14				
	11220026							20		16			
	7104-014										22		
	7312-031											21	
Population	Unrelated										25	19	11

# ➤ Optimisation du coût de génotypage

Exemple avec l'imputation



Impact de la densité de marqueurs

# ➤ Optimisation du coût de génotypage

Exemple avec l'imputation

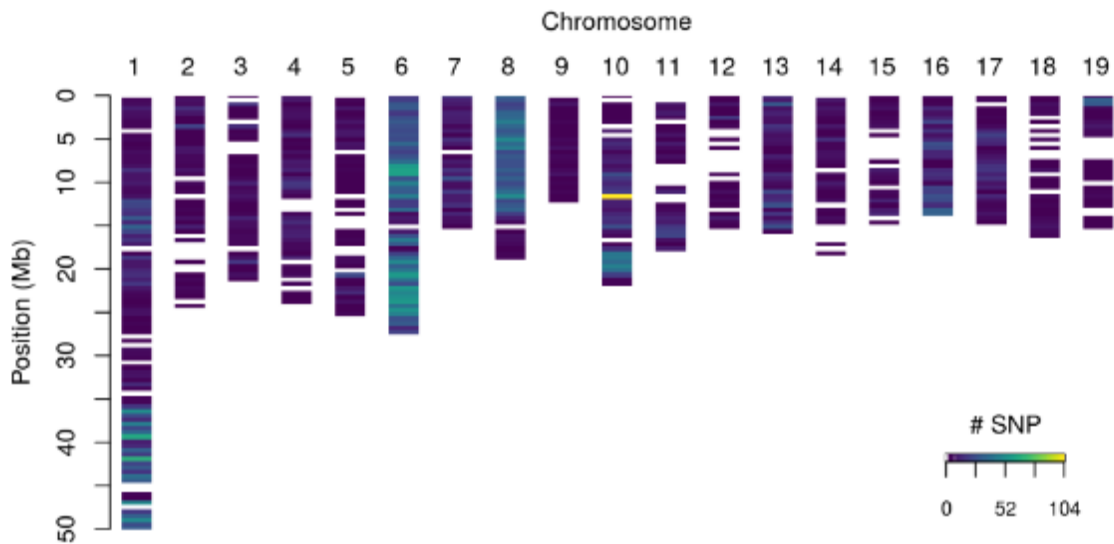
L'imputation a permis de passer de 7540 SNPs à 1.4 millions

Meilleure couverture du génome

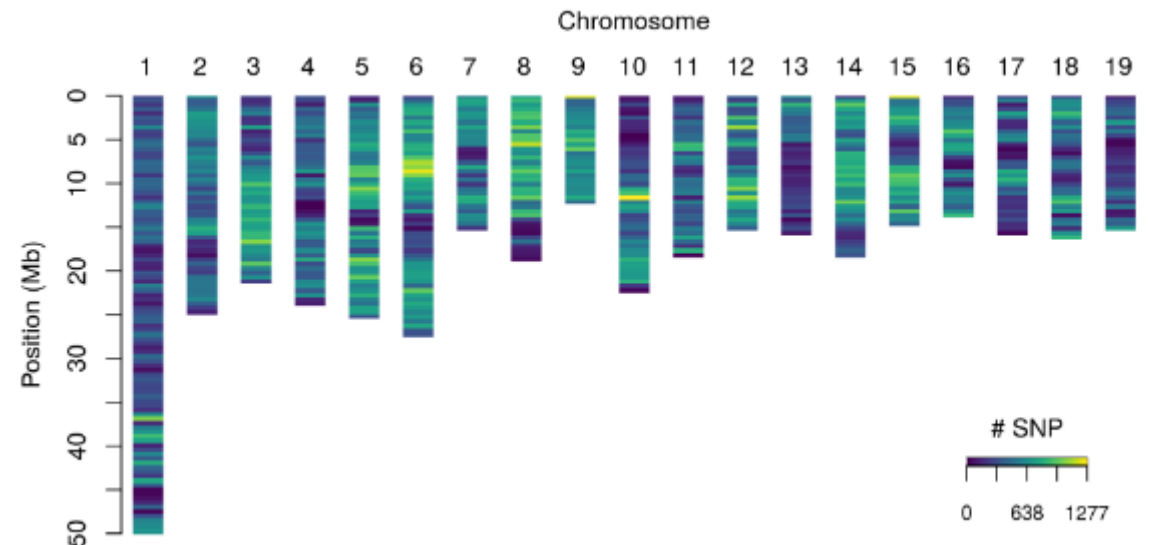
Plus homogène que précédemment

Coût réduit

Puce 7540 SNPs



Après imputation



## ➤ Conclusions

Une multitude de méthodes avec des avantages et inconvénients

Adapter la méthode à la question posée et aux connaissances de l'espèce d'intérêt

Génotypage de plus en plus spécialisé à sous traiter tout ou partie (Macrogen, Keygene, LGC genomics...)





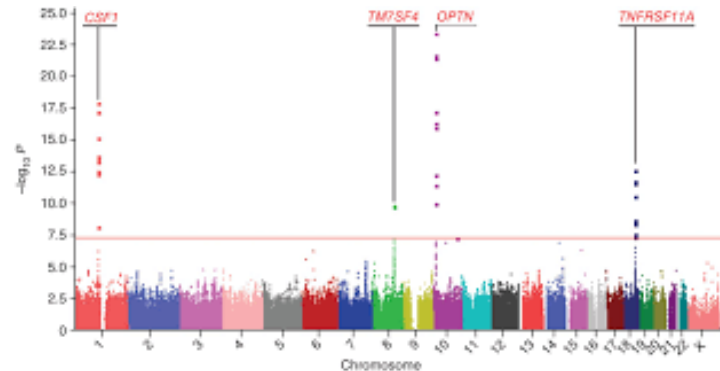
## ➤ Conclusions

Dans l'avenir séquençage complet systématique??

Outils de bio-informatique cruciaux, problème de stockage

Changement des outils d'analyse des données de génotypage et de phénotypage maintenant en très grand nombre

Des outils impressionnants pour mieux comprendre la diversité du monde vivants !!!!



INRAE

➤ Utilisation des marqueurs en amélioration  
des plantes

Partie 2

## ➤ Distinction entre variété/ individus

Dans quels cas

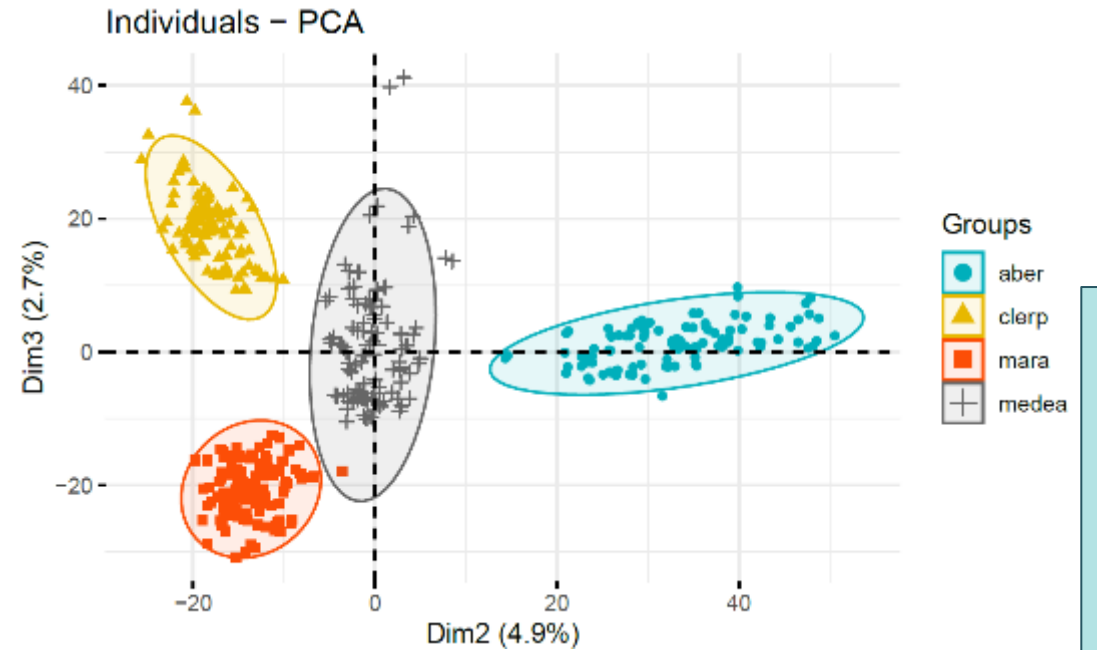
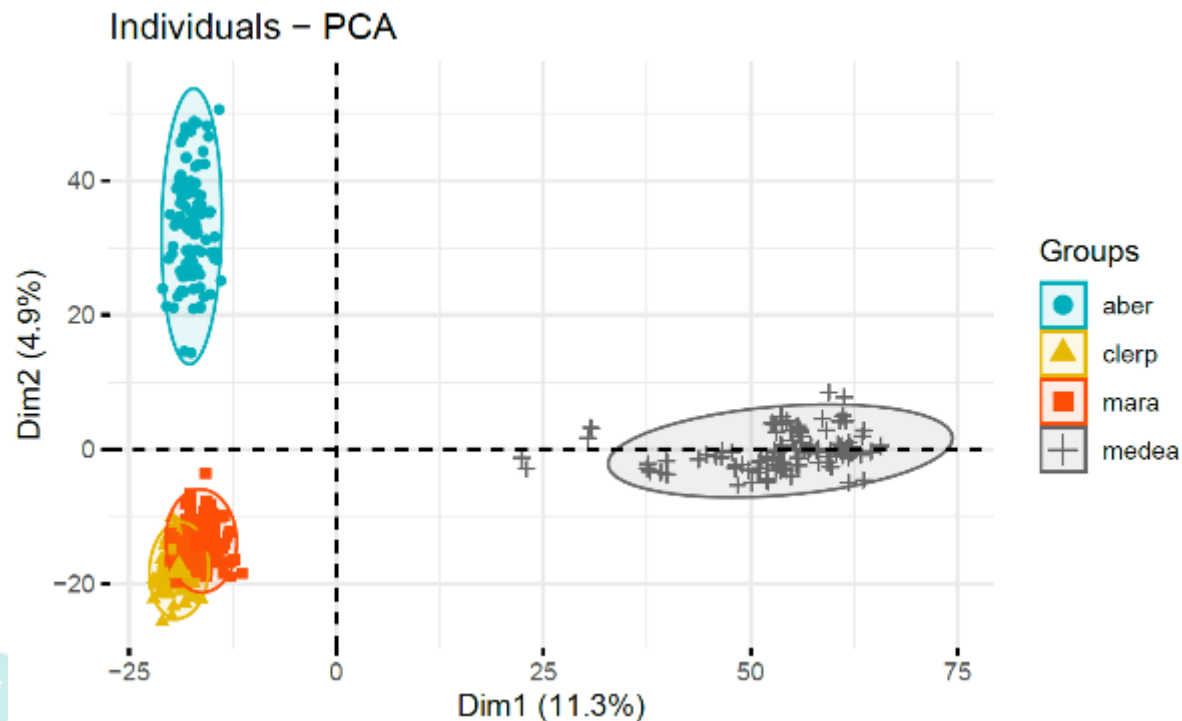
- Pour gérer le matériel du sélectionneur :
  - Problème d'échantillonnage
  - Perte d'identifiant ou dérive entre le phénotype observé et attendu
- Vérification pré-inscription
  - Pour vérifier que la nouvelle variété est suffisamment différente des parents
    - si on utilise une variété du sélectionneur concurrent comme parent par exemple
- Vérification post-inscription
  - Vérification des lots de multiplication d'une variété
  - Contrôle de la dérive génétique pour les variétés population

# ➤ Distinction entre variété/ individus

exemple

Génotypage de plusieurs individus appartenant à des variétés.

On cherche à savoir à quel variété appartient l'échantillon



Projet Invite  
Dactyle  
Génotypage via capture

# ➤ Test de paternité

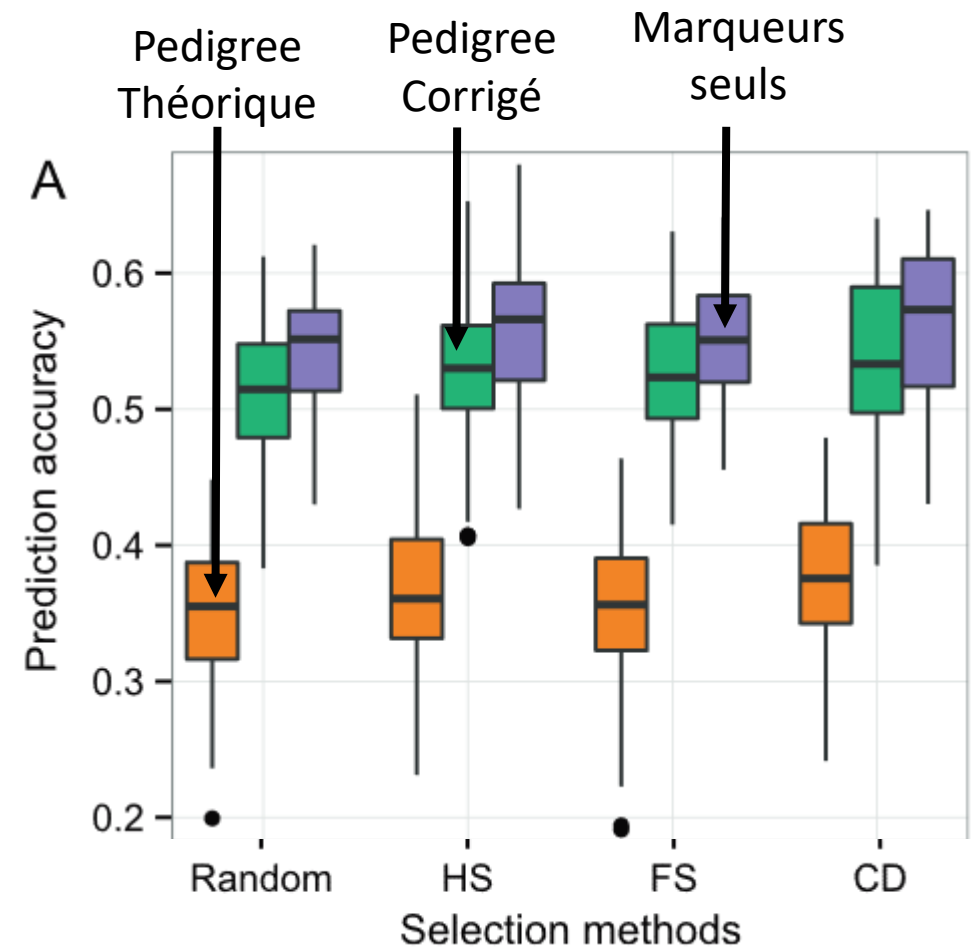
Dans quels cas

Pour faciliter le travail du sélectionneur :

- Croisement contrôlés en cage avec un mélange de pollen (On retrouve le père avec les marqueurs)
- Correction du pedigree et vérification de la présence de pollution génétique.


Exemple Pin maritime :

- Différent méthodes de sélection
- Prédiction des valeurs des descendants selon le pedigree ou les marqueurs
- Correction du pedigree à l'aide des marqueurs

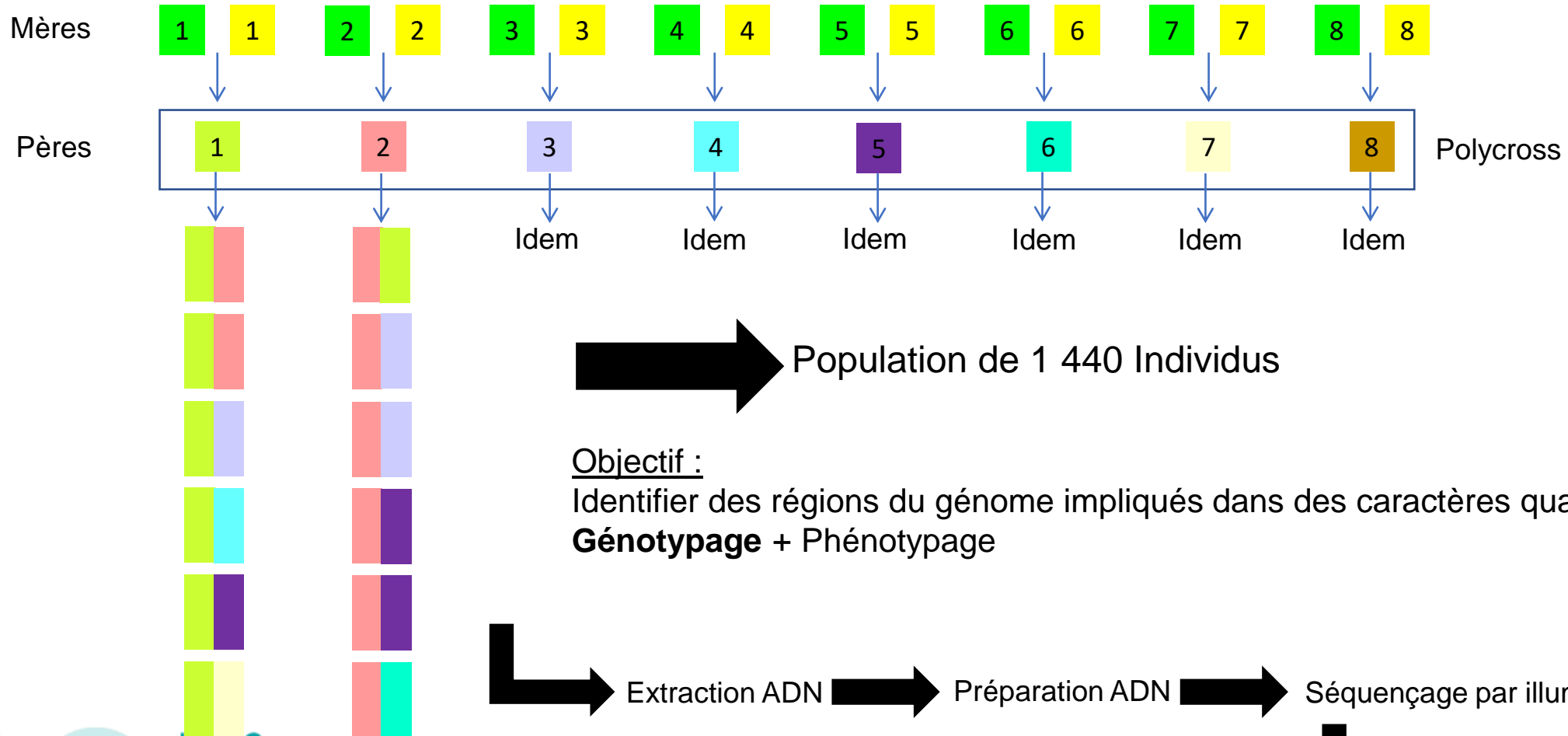


# ➤ Test de paternité

Autre exemple

Tempéré élite 

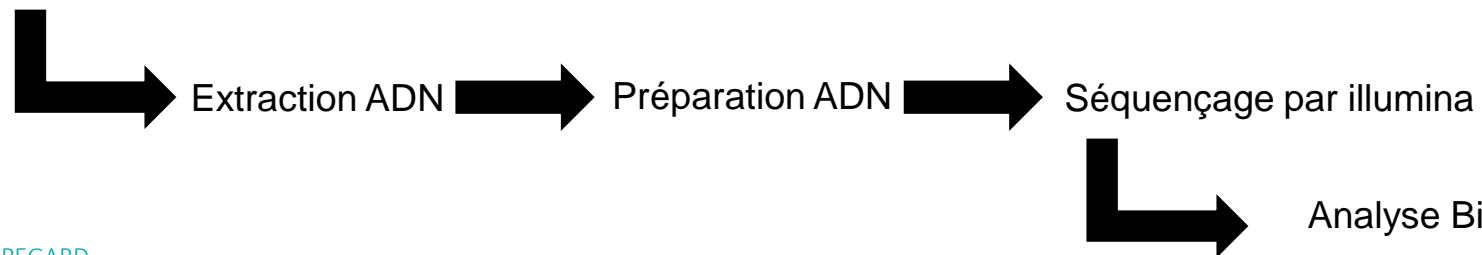
Méditerranéen résistant 



Objectif :

Identifier des régions du génome impliqués dans des caractères quantitatifs (QTL) :

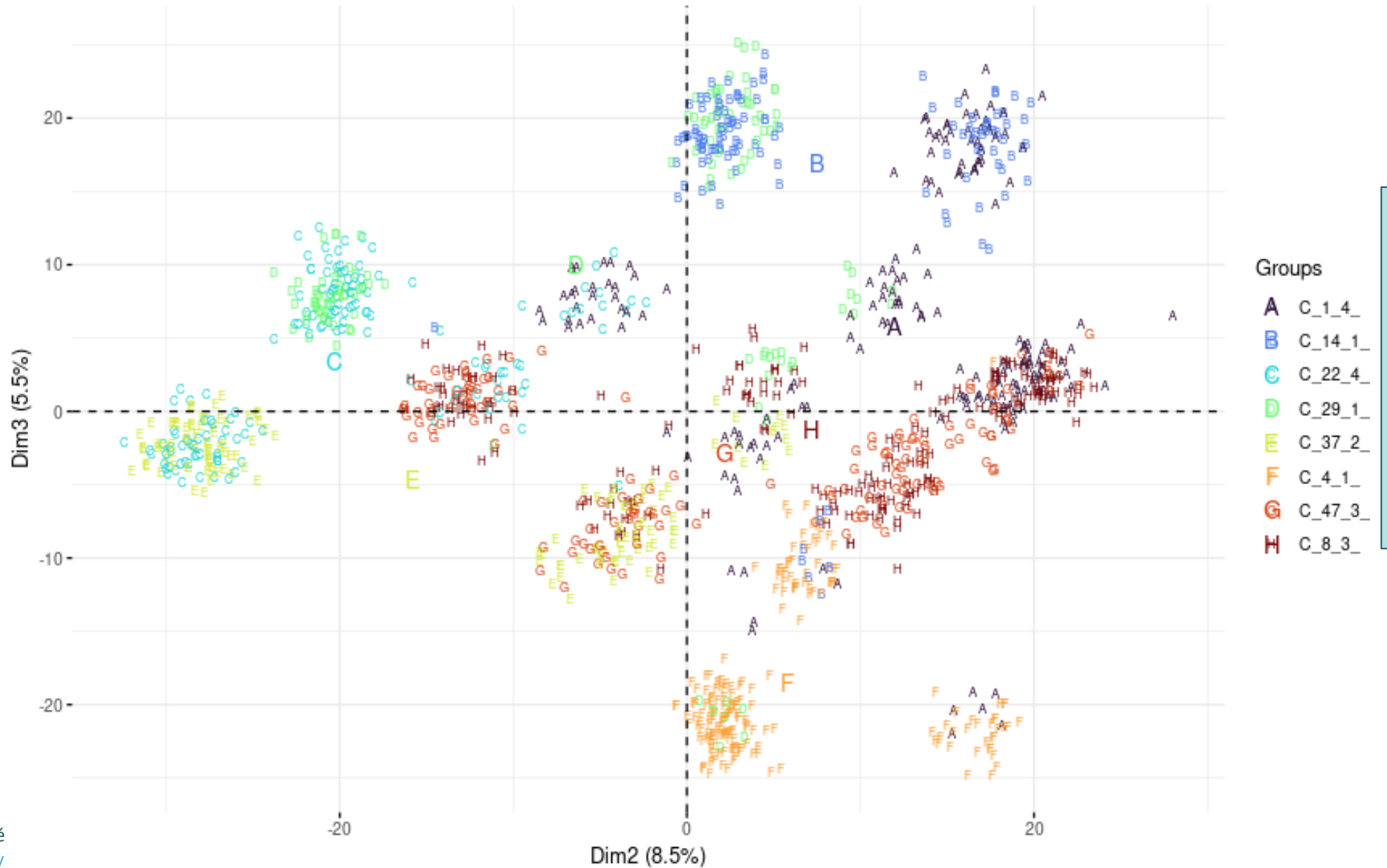
**Génotypage** + Phénotypage



# ➤ Test de paternité

Autre exemple

Individuals - PCA



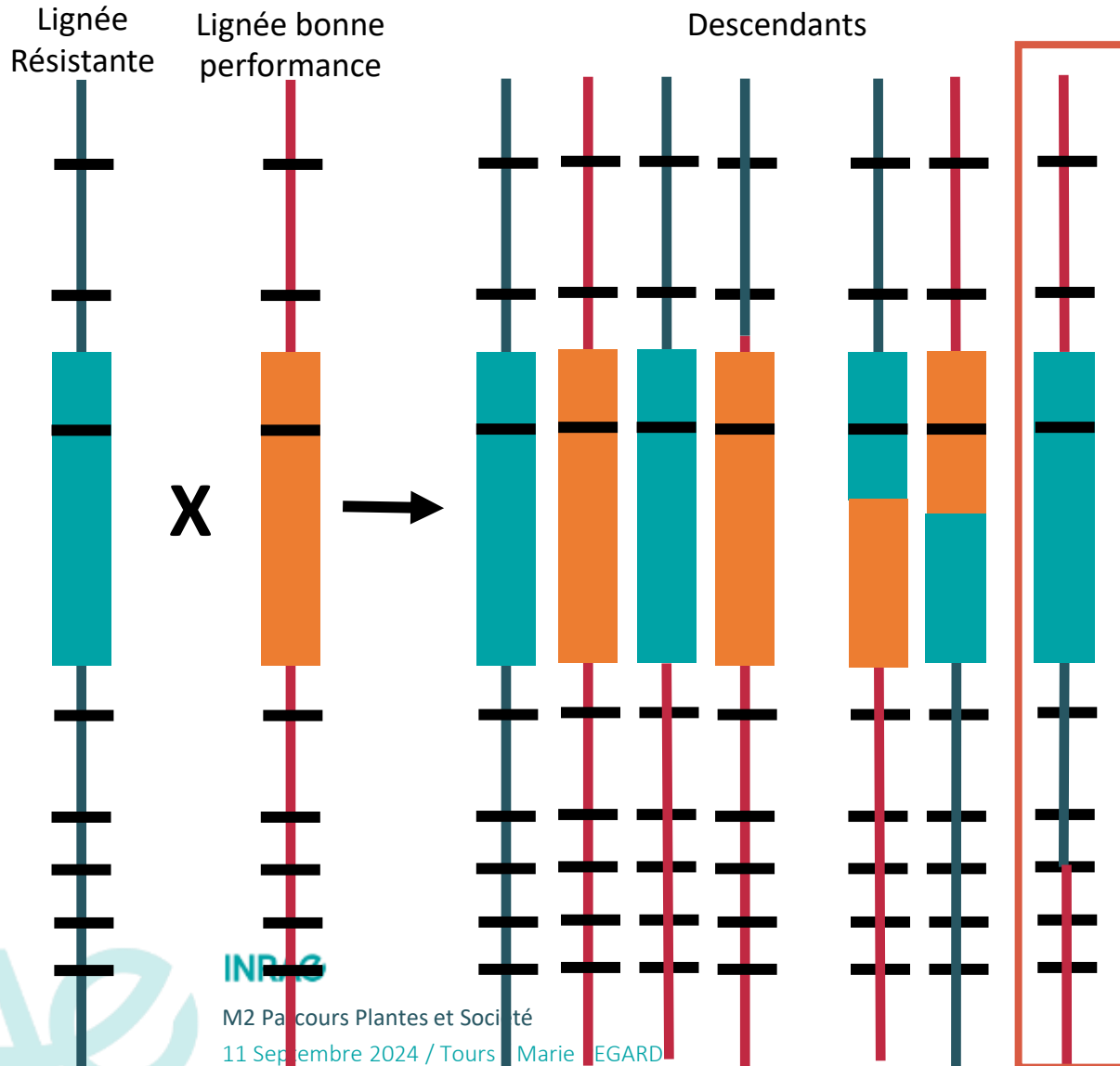
ACP des descendants classés par famille de demi-frère

Partie II



# ➤ Suivi introgression / backcross

Comment utiliser les marqueurs : trier les descendants

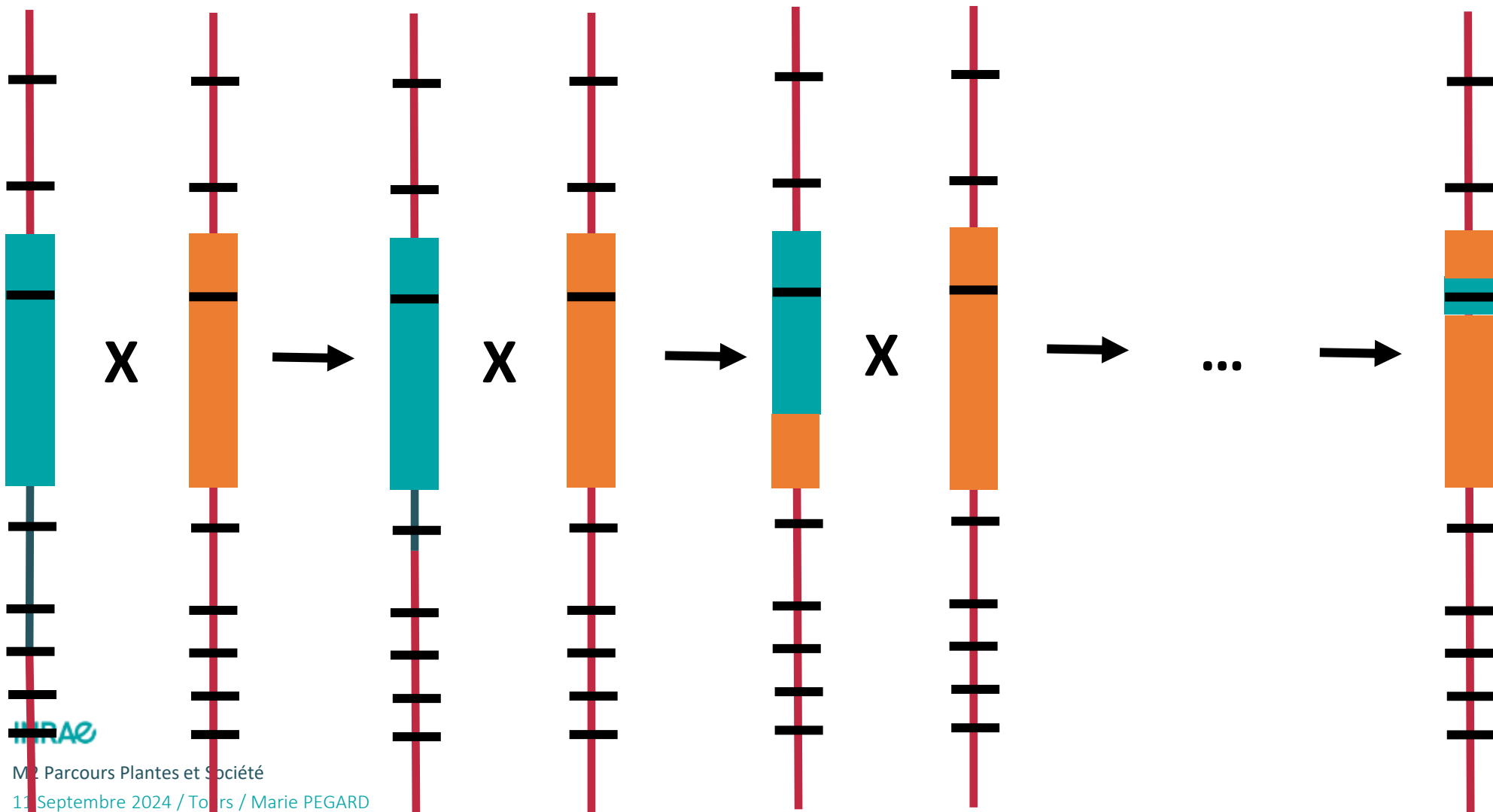


Parmi tous les descendants celui avec le gène de résistance et la plus grande portion du génome de la lignée Performante est choisi



## ➤ Suivi introgression / backcross

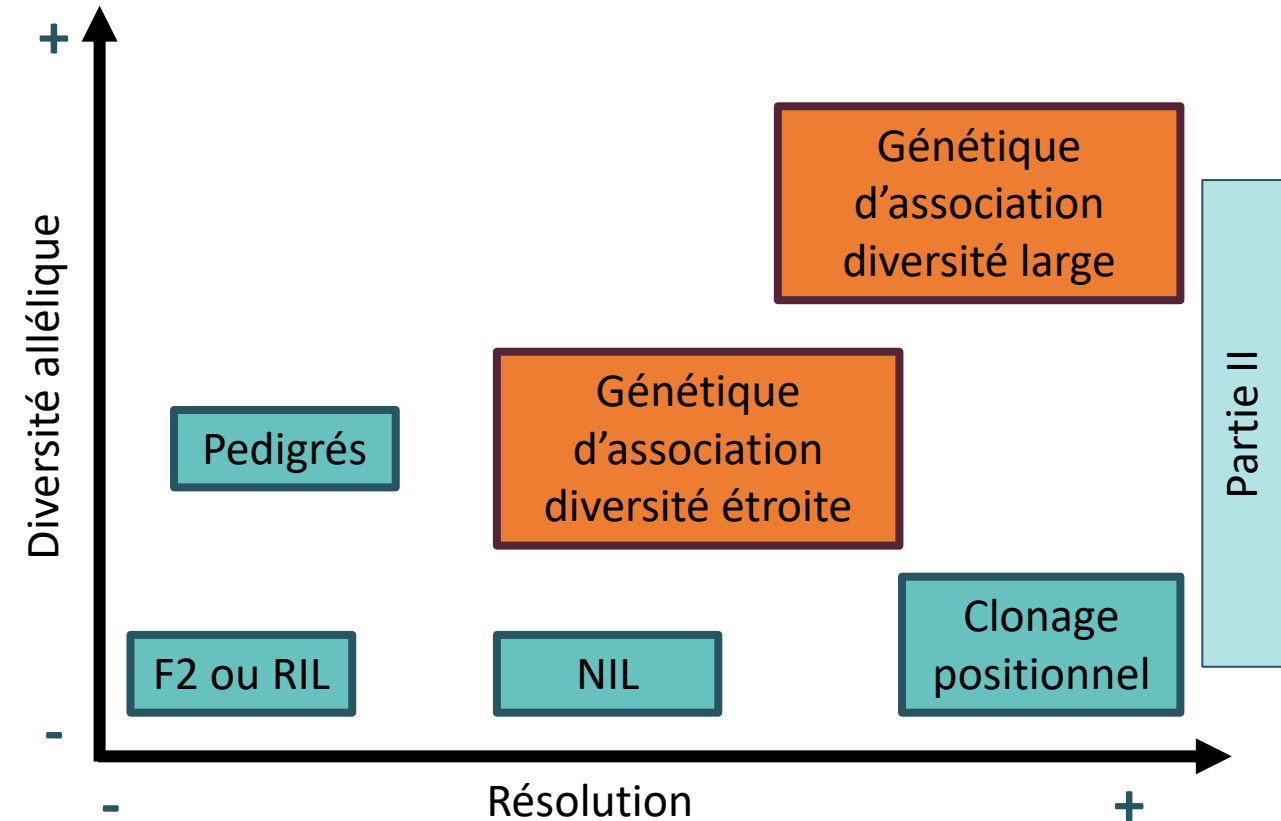
Les marqueurs permettent de suivre l'introgression à chaque étape et de sélectionner les descendants ayant le gène de résistance et un plus grande portion du génome élite



# ➤ Détection QTL

Deux possibilités

- Via une analyse de liaison
  - Population de cartographie
  - Croisement contrôlé
  - Parent et apparentement connu
  - Intervalle de confiance large autour de l'allèle = Résolution faible
- Via la génétique d'association (GWAS)
  - Population naturelle
  - Croisement non contrôlé
  - Parent et apparentement inconnu
  - Intervalle de confiance faible autour de l'allèle = Résolution forte

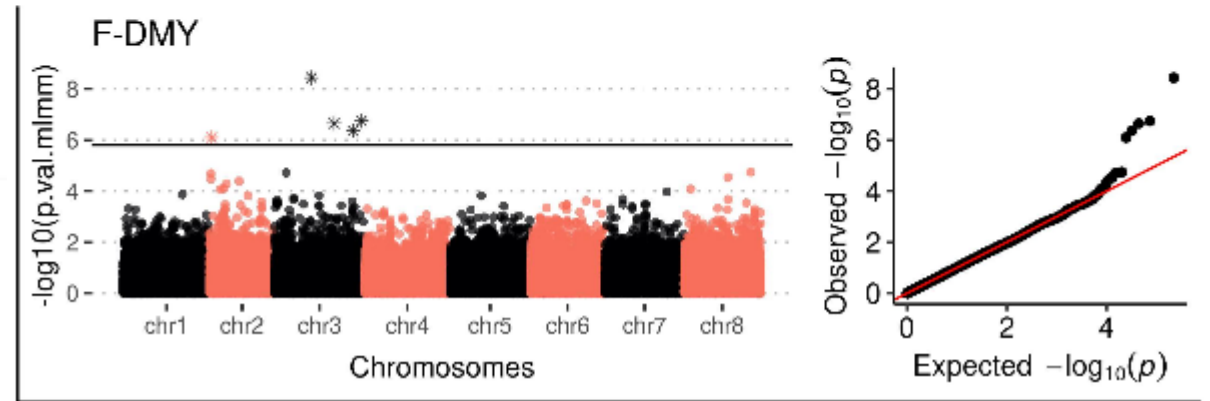
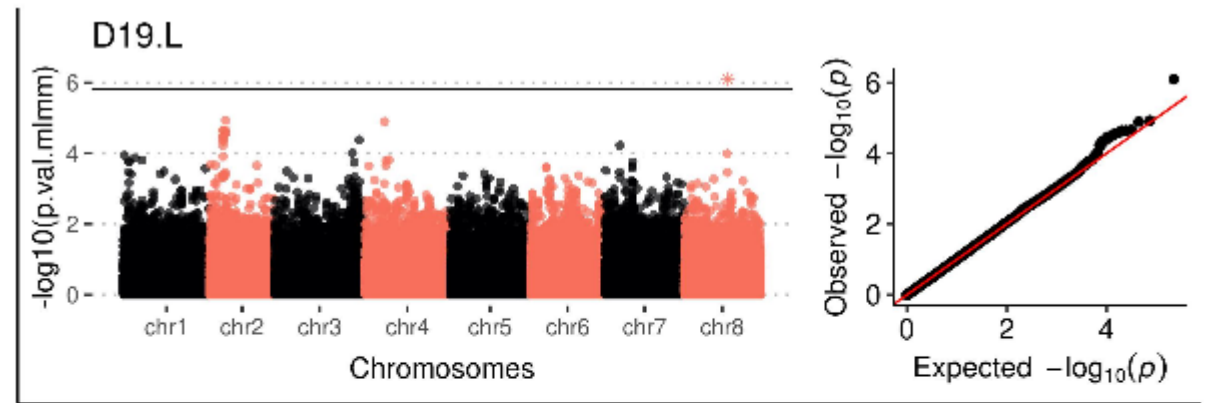


# ➤ Détection QTL

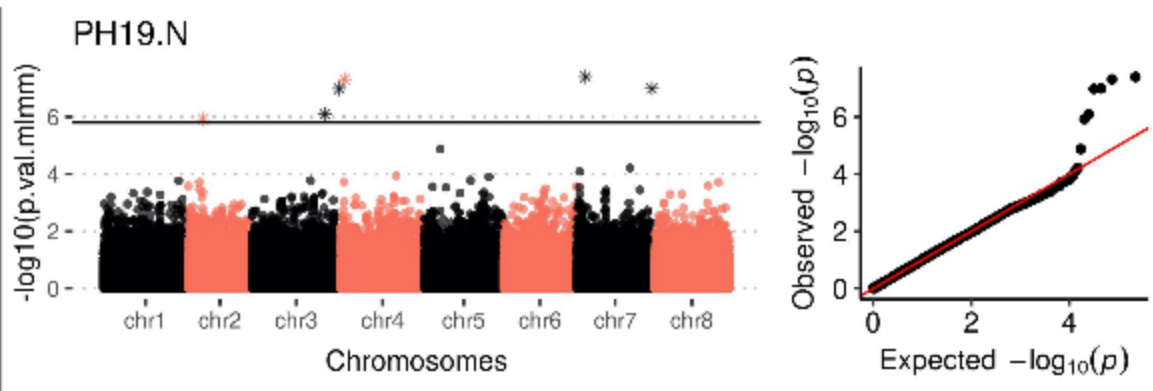
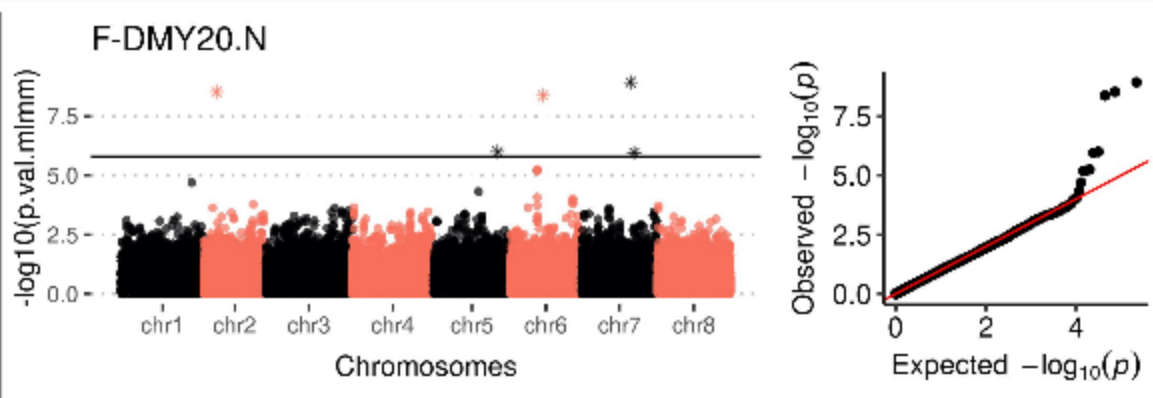
## Exemple de GWAS

400 populations de Luzerne (EUCLEG)

- Génotypage GBS (118 K SNP)
- Phénotypage de la dormance automnale



Partie II



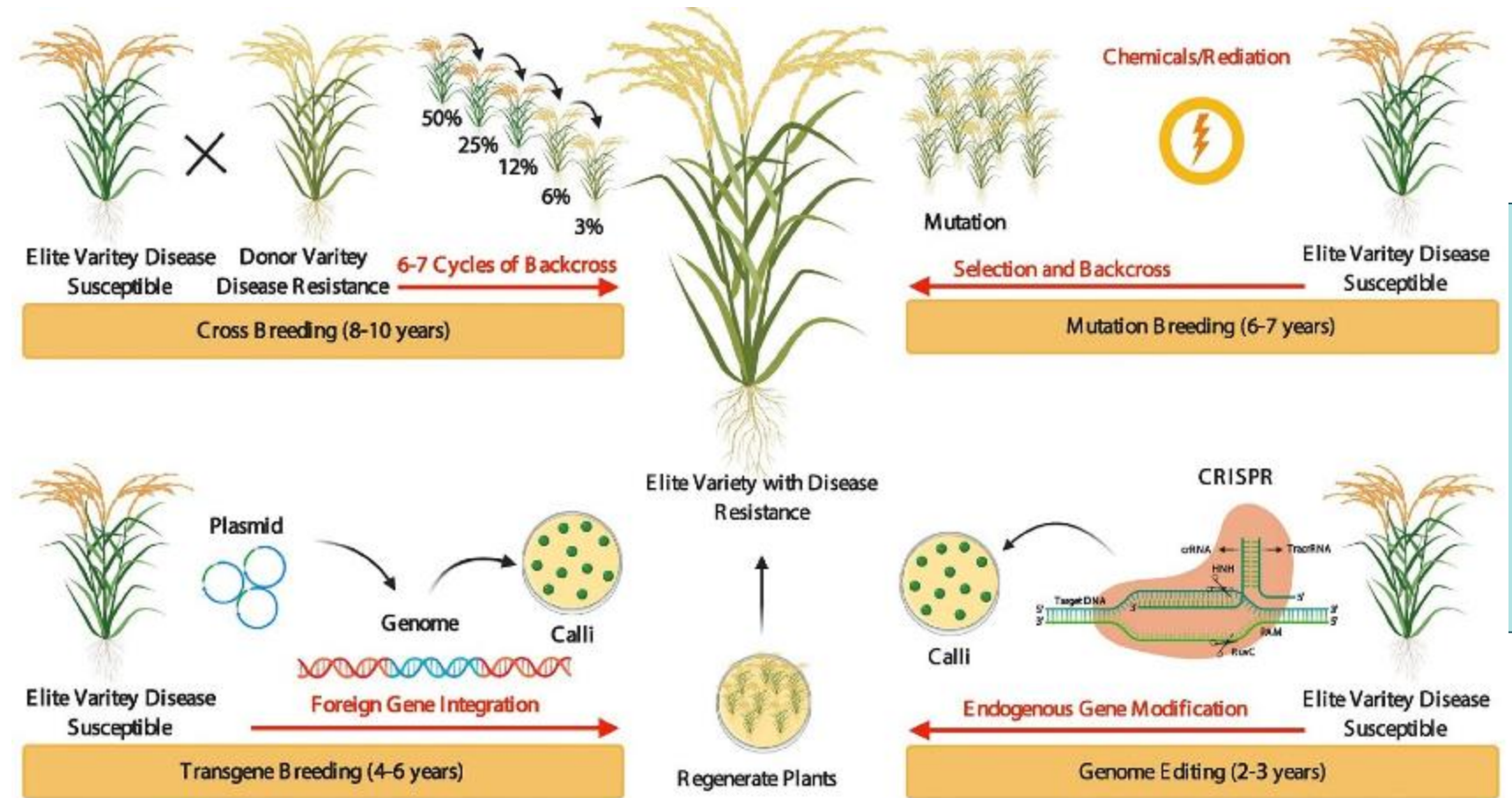
L'ensemble des QTL se trouve dans ou a moins de 5000 bp d'un gène

# ➤ Détection QTL

## Débouchés

Différents débouchés

- Introgression (BackCross)
- Pyramidage : BackCross de plusieurs gènes)
- Transformation ciblée (CRISPR/Cas9)
- Sélection Assisté par marqueur (MAS)



Wang, Y.; Zafar, N.; Ali, Q.; Manghwar, H.; Wang, G.; Yu, L.; Ding, X.; Ding, F.; Hong, N.; Wang, G.; et al. CRISPR/Cas Genome Editing Technologies for Plant Improvement against Biotic and Abiotic Stresses: Advances, Limitations, and Future Perspectives. *Cells* **2022**, *11*, 3928. <https://doi.org/10.3390/cells11233928>

## ➤ Rappels sur la sélection génomique...

Qu'est-ce que la sélection génomique?

Meuwissen, Hayes et Goddard, 2001 :

"Estimer les effets de ~50 000 marqueurs haplotypes simultanément à partir d'un nombre limité de données phénotypiques".

"La sélection sur les valeurs génétiques prédites par les marqueurs pourrait augmenter substantiellement le taux de gain génétique chez les animaux et les plantes".

Genetics 157: 1819–1829 (April 2001)

### **Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps**

**T. H. E. Meuwissen,\* B. J. Hayes<sup>†</sup> and M. E. Goddard<sup>†,‡</sup>**

*\*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, <sup>†</sup>Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and <sup>‡</sup>Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia*

Manuscript received August 17, 2000  
Accepted for publication January 17, 2001



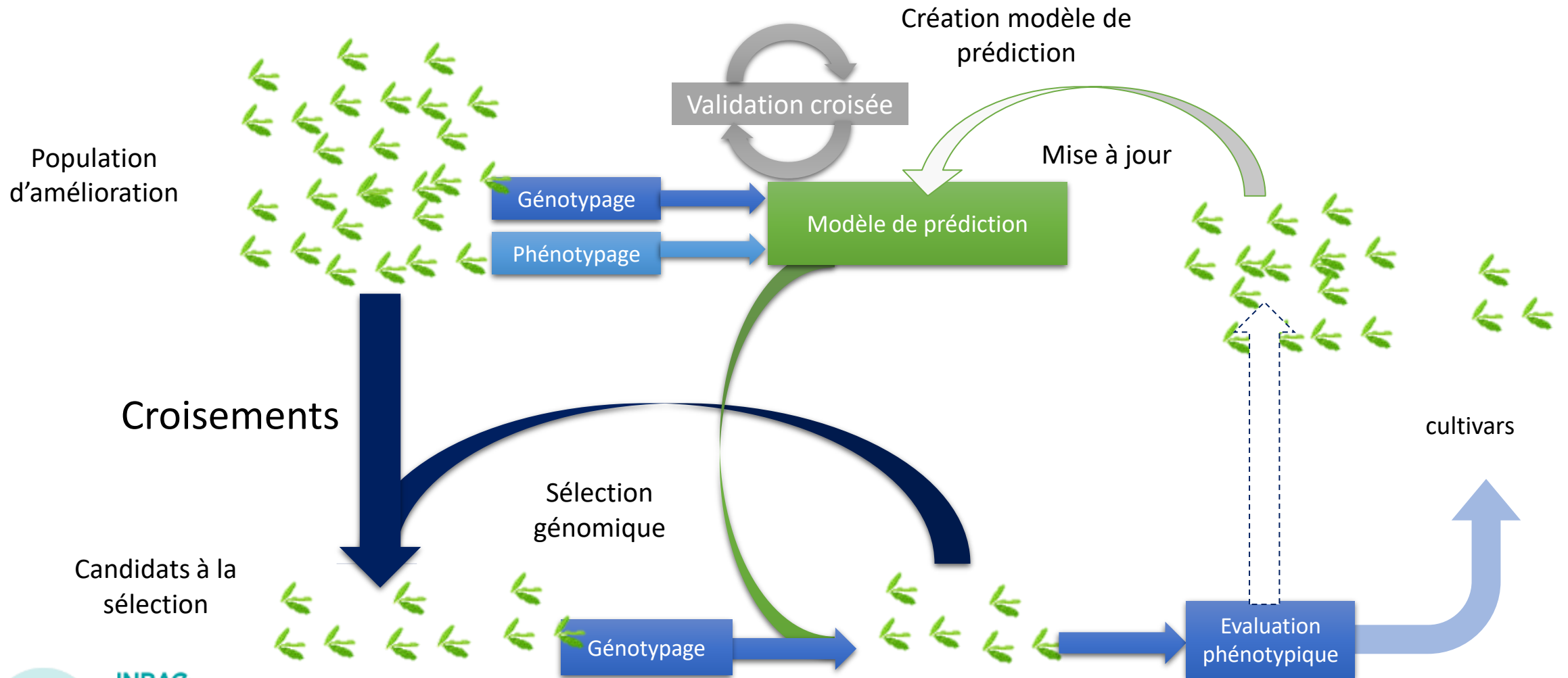
**INRAE**

M2 Parcours Plantes et Société

11 Septembre 2024 / Tours / Marie PEGARD

# ➤ Rappels sur la sélection génomique...

Qu'est-ce que la sélection génomique?



## ➤ Rappels sur la sélection génomique...

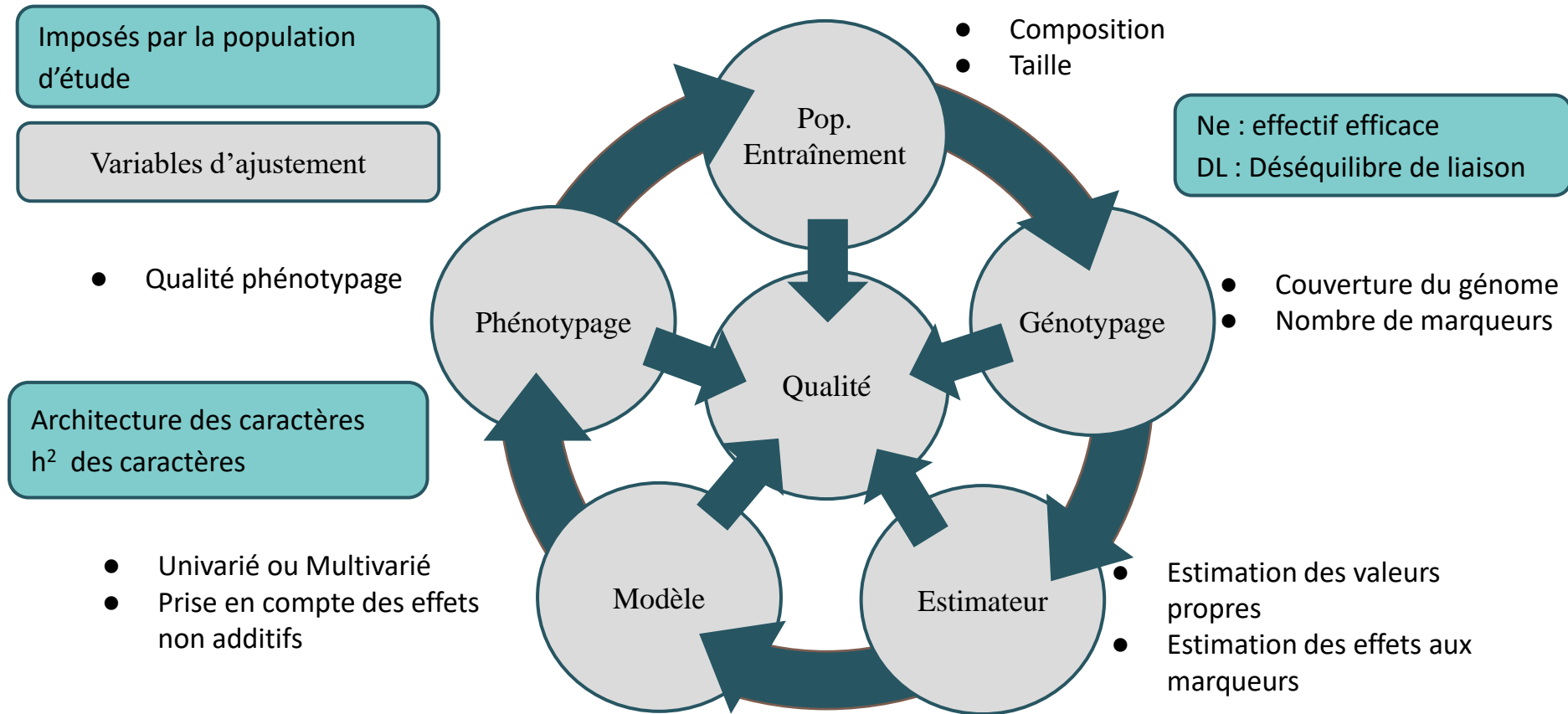
Les attentes vis à vis de la SG



The diagram shows the equation for genetic gain:  $\text{Gain génétique} = \frac{i r \sigma_A}{L}$ . A large green arrow points upwards to the left of the text. Three green arrows point upwards from the variables  $i$ ,  $r$ , and  $\sigma_A$  in the numerator. One orange arrow points downwards from the variable  $L$  in the denominator.

$$\text{Gain génétique} = \frac{i r \sigma_A}{L}$$

# ➤ Quels facteurs impactent la qualité de la prédiction génomique?

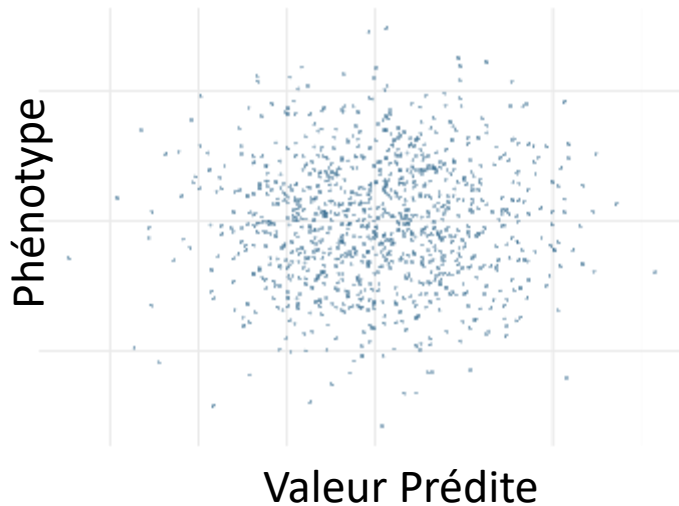




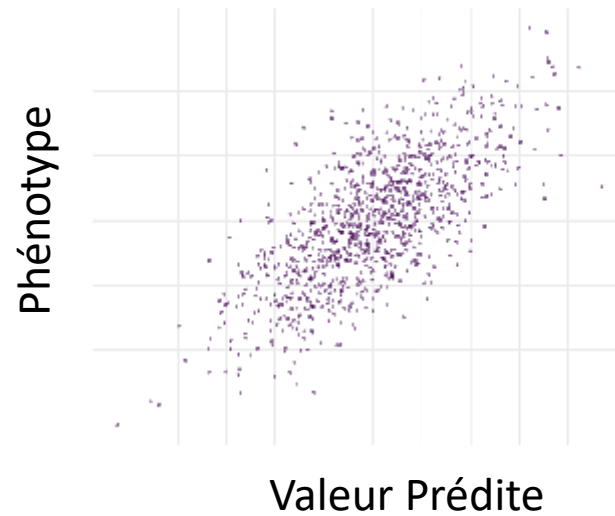
# ➤ Quels facteurs impactent la qualité de la prédiction génomique?

Mesure de la qualité de prédiction

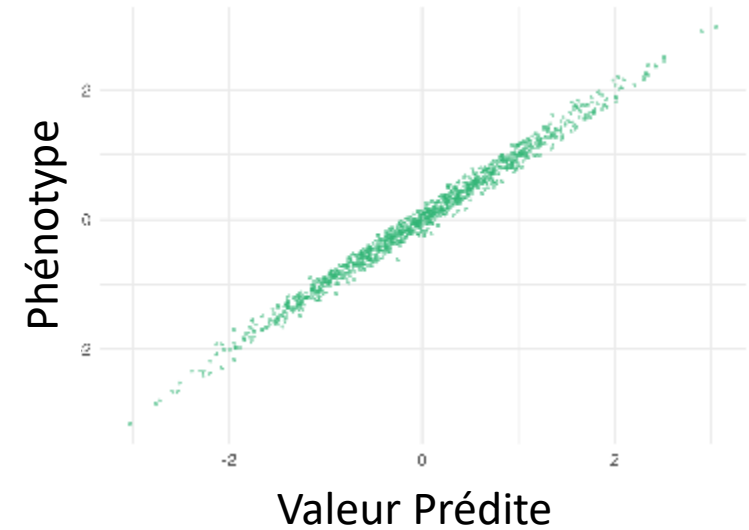
**Qualité de prédiction** : Corrélation de Pearson



Faible



Moyenne



Forte

# ➤ Quels facteurs impactent la qualité de la prédiction génomique?

Mesure de la qualité de prédiction

**Qualité de prédiction** : Corrélation de Pearson

**Précision** :  $\frac{\text{Corrélation}}{\text{Héritabilité}}$

Héritabilité : Variance génétique expliquée par le modèle

Héritabilité au sens strict : Variance des effets additifs expliquée par le modèle

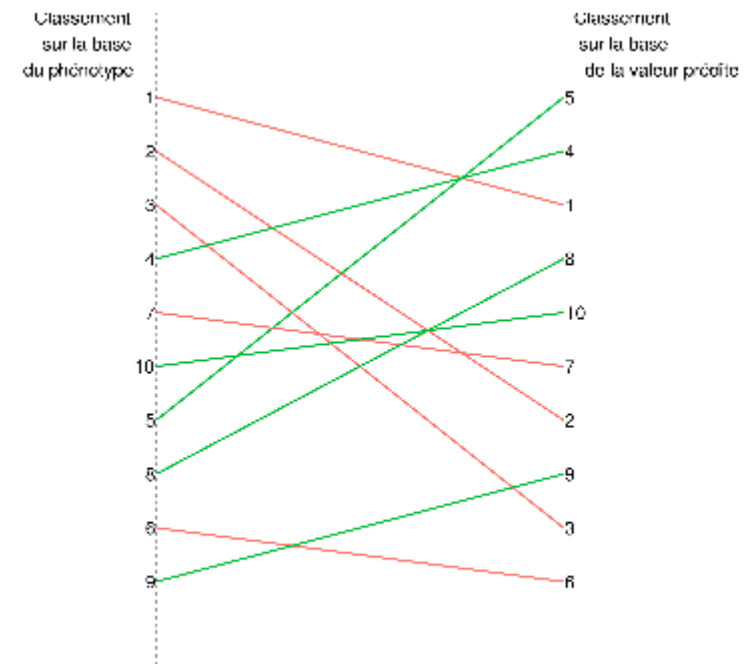
# ➤ Quels facteurs impactent la qualité de la prédiction génomique?

Mesure de la qualité de prédiction

**Qualité de prédiction** : Corrélation de Pearson

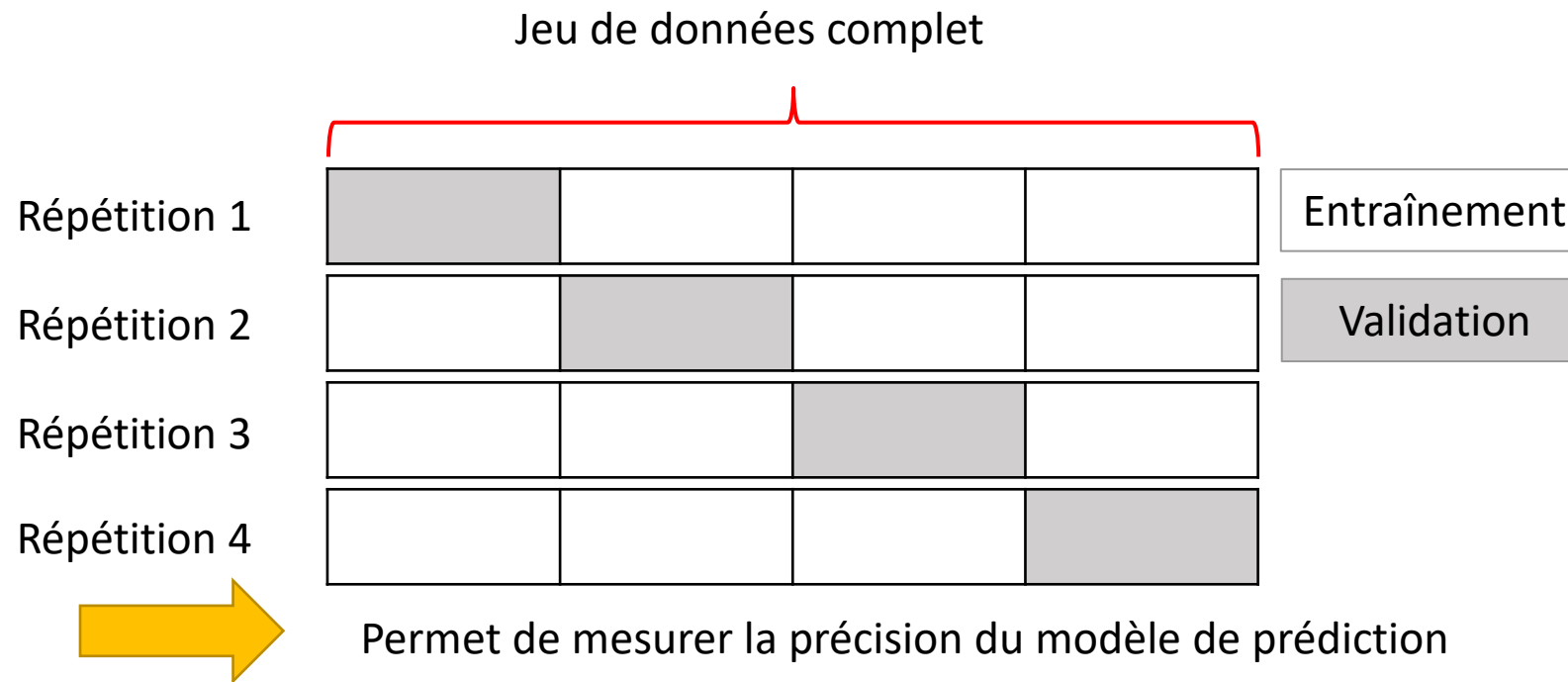
**Précision** :  $\frac{\text{Corrélation}}{\text{Héritabilité}}$

**Classement** : Corrélation de Spearman



# ➤ Quels facteurs impactent la qualité de la prédiction génomique?

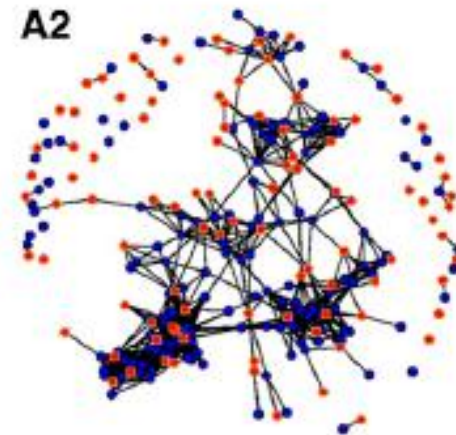
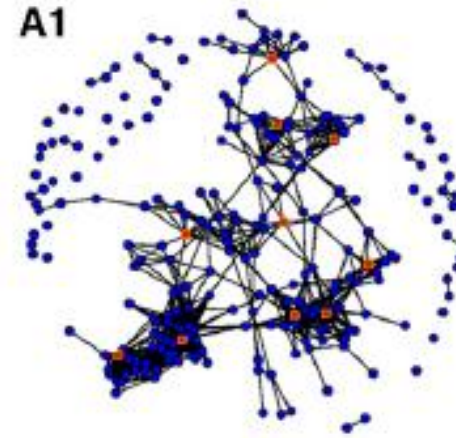
Mesure de la qualité de prédiction



# ➤ Quels facteurs impactent la qualité de la prédiction génomique?

Taille et composition de la population d'entraînement

	Groupe 1		Groupe 2		Groupe 3		Groupe 4	
T50.Individus								
T25.Individus								
T50.Groupe								
T25.Groupe								



CDmean (Rincent et al, 2012 Genetics)

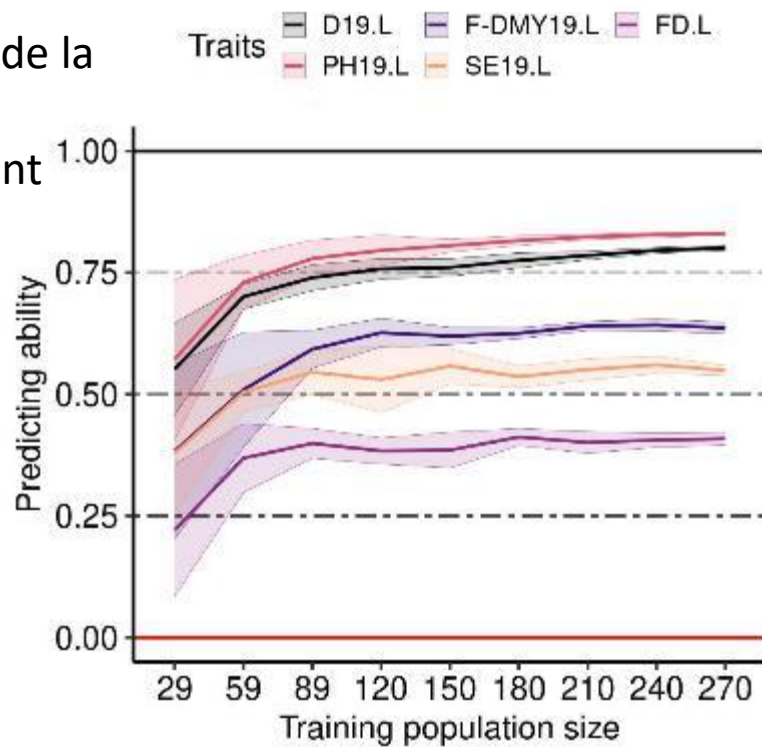
# ➤ Optimisation schéma de sélection

## Exemple Projet Eucleg

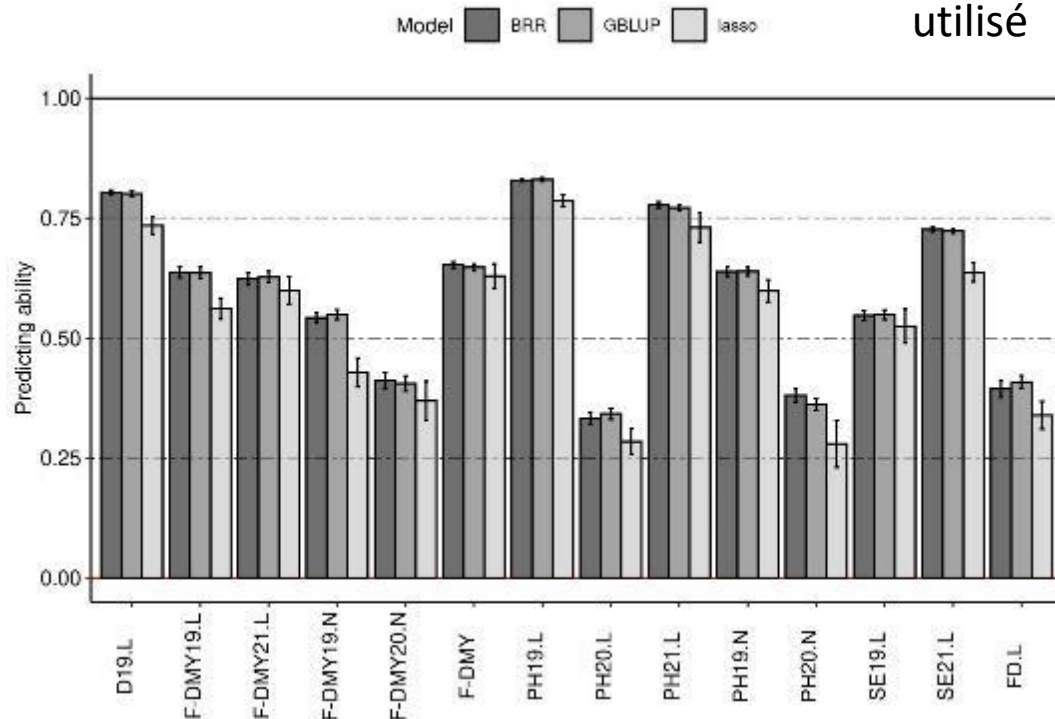
400 populations de Luzerne (EUCLEG)

- Génotypage GBS (118 K SNP)
- Phénotypage de la dormance automnale

Effet de la taille de la population d'entraînement



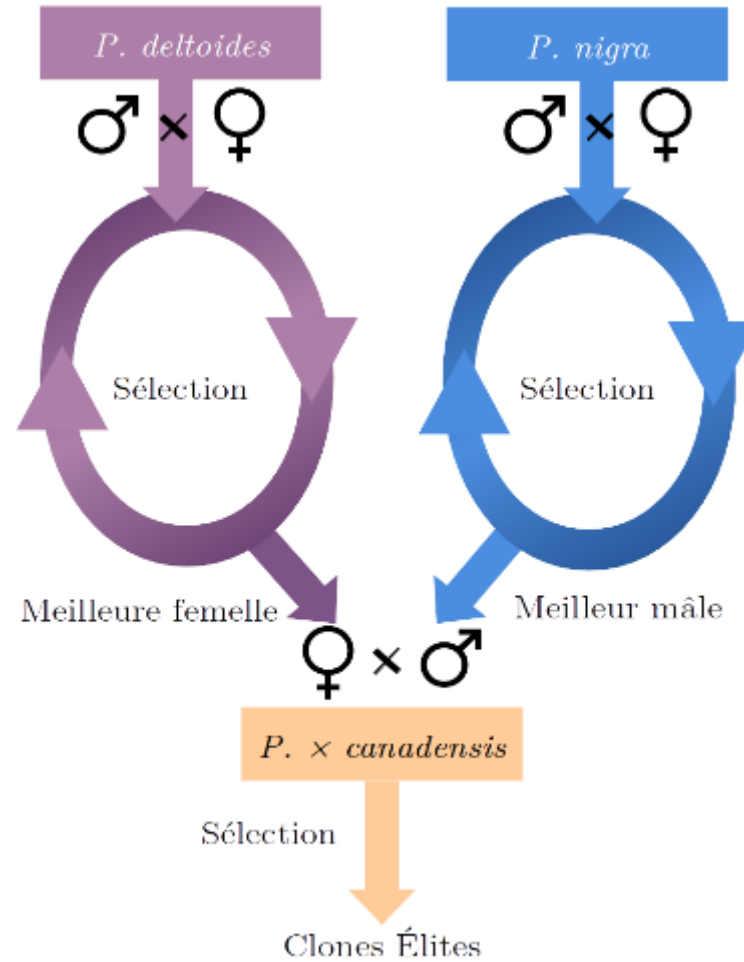
Effet du modèle utilisé



# ➤ Optimisation schéma de sélection

Programme d'amélioration français (GIS peuplier)

Fécondation fleur *P. deltoides* ♀



Fleur ♂ *P. nigra* & récolte pollen



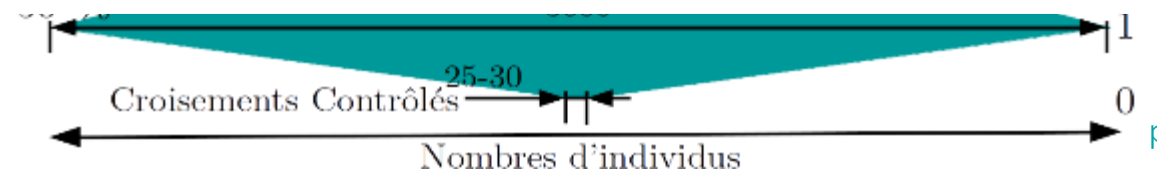
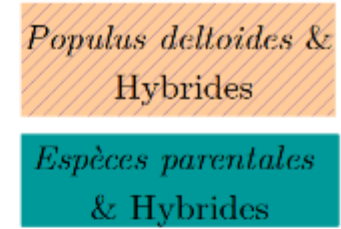
Partie II

# ➤ Optimisation schéma de sélection

Programme d'amélioration actuel

Taux de sélection

Années



Partie II



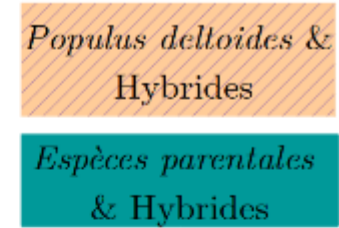


# ➤ Optimisation schéma de sélection

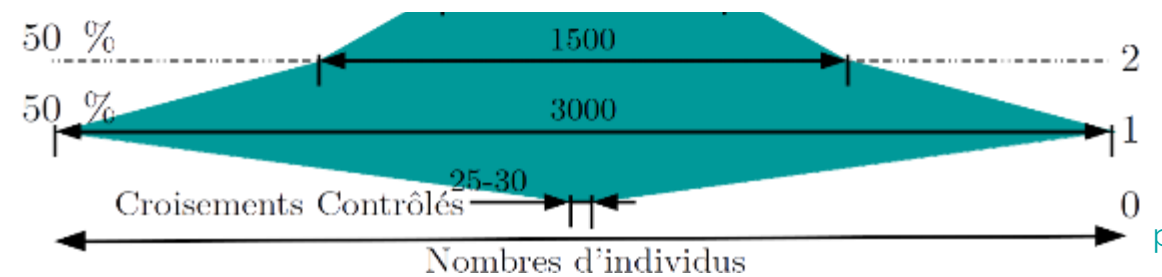
Programme d'amélioration actuel

Taux de sélection

Années



Multiplication végétative



Partie II

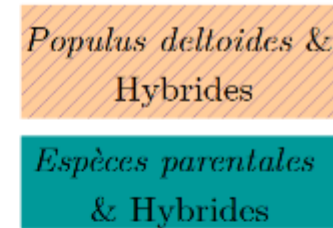


# Optimisation schéma de sélection

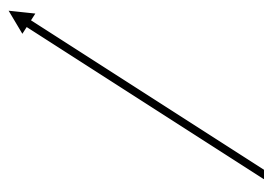
Taux de sélection

Années

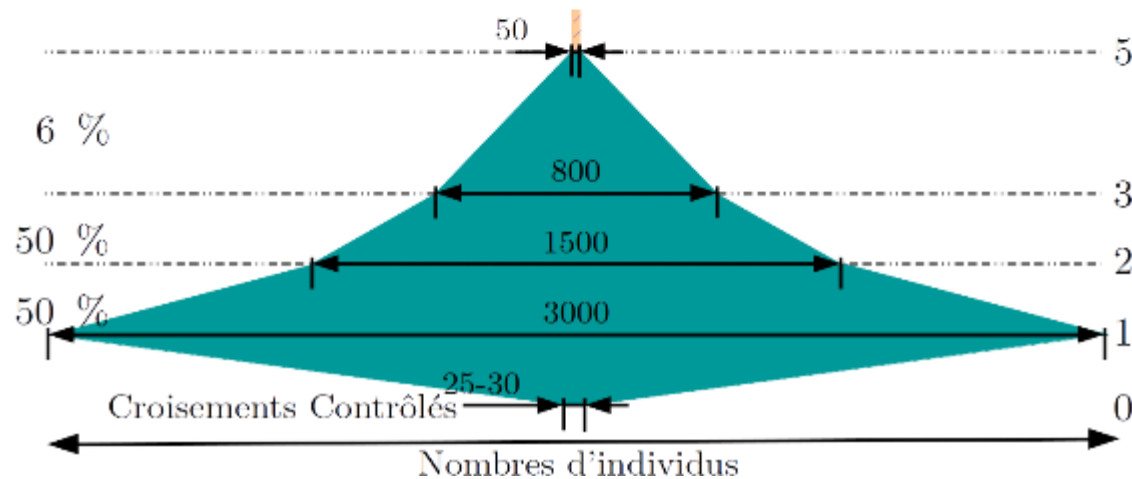
Programme d'amélioration actuel



Pépinière :  
Croissance  
Phénologie  
angle de branche  
résistance à la rouille



Multiplication  
végétative



Partie II



INRAE

M2 Parcours Plantes et Société

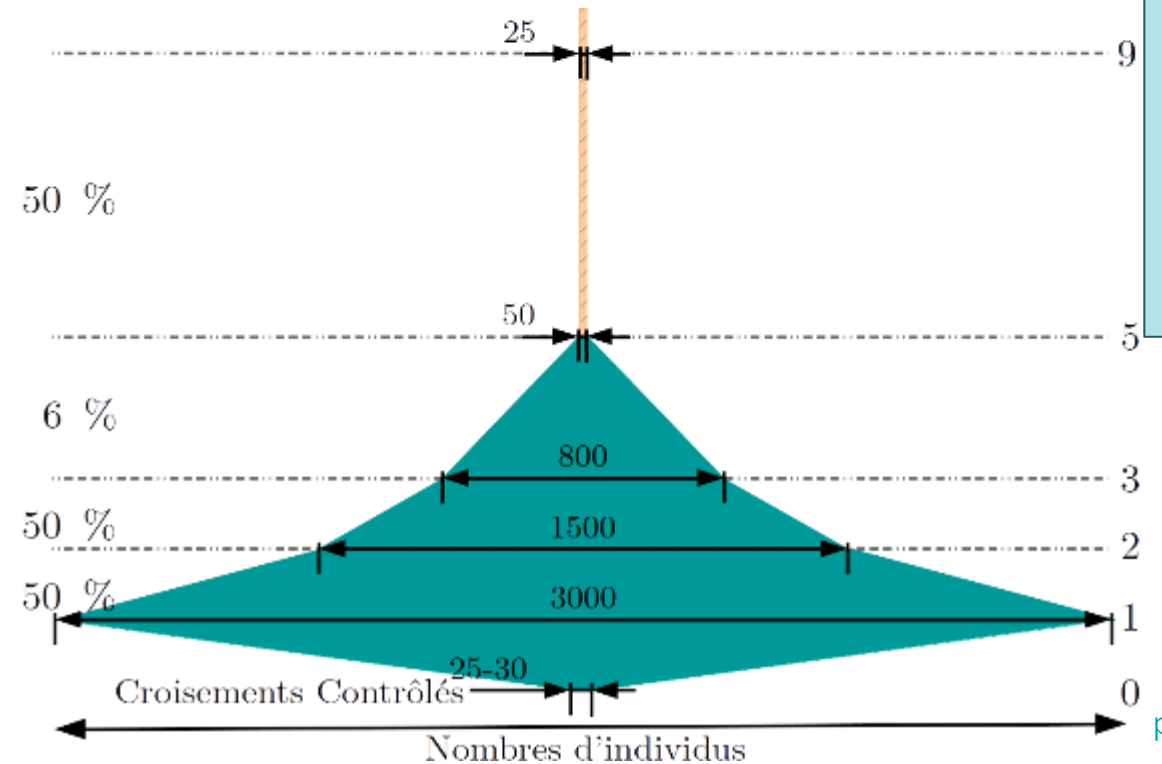
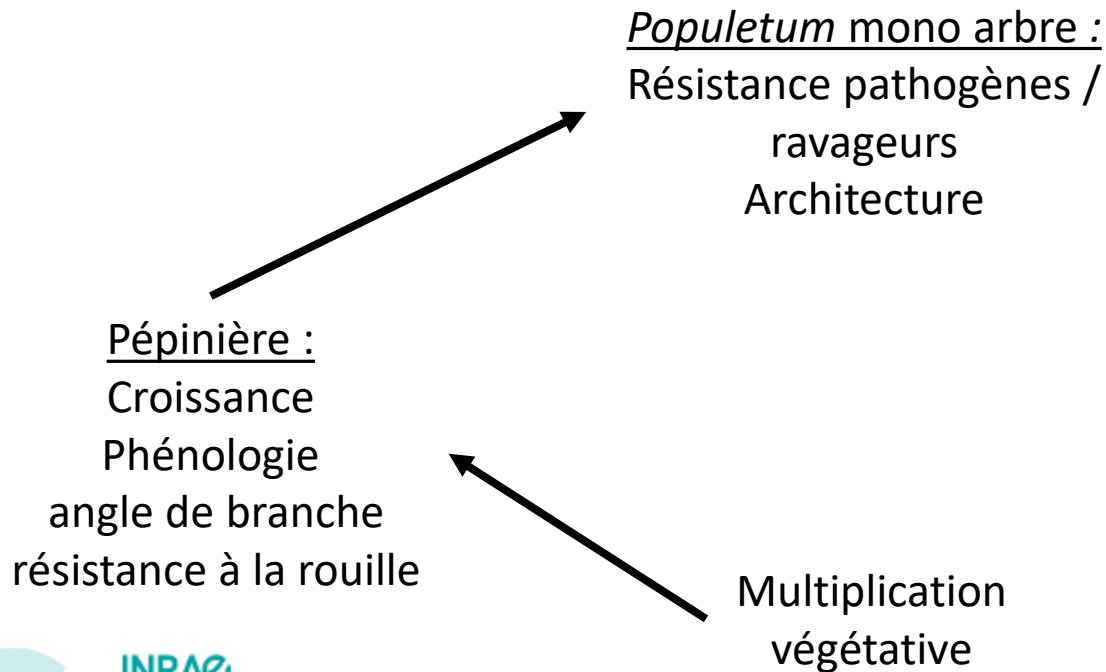
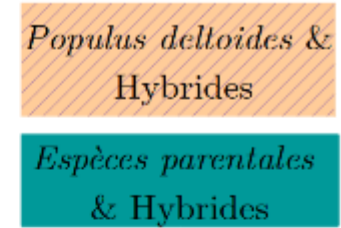
11 Septembre 2024 / Tours / Marie PEGARD

# Optimisation schéma de sélection

Programme d'amélioration actuel

Taux de sélection

Années



Partie II



# ➤ Optimisation schéma de sélection

Programme d'amélioration actuel

Populetum pluri-arbre :

- Croissance
- Architecture
- qualité et propriété du bois
- Sexe
- interaction GxE

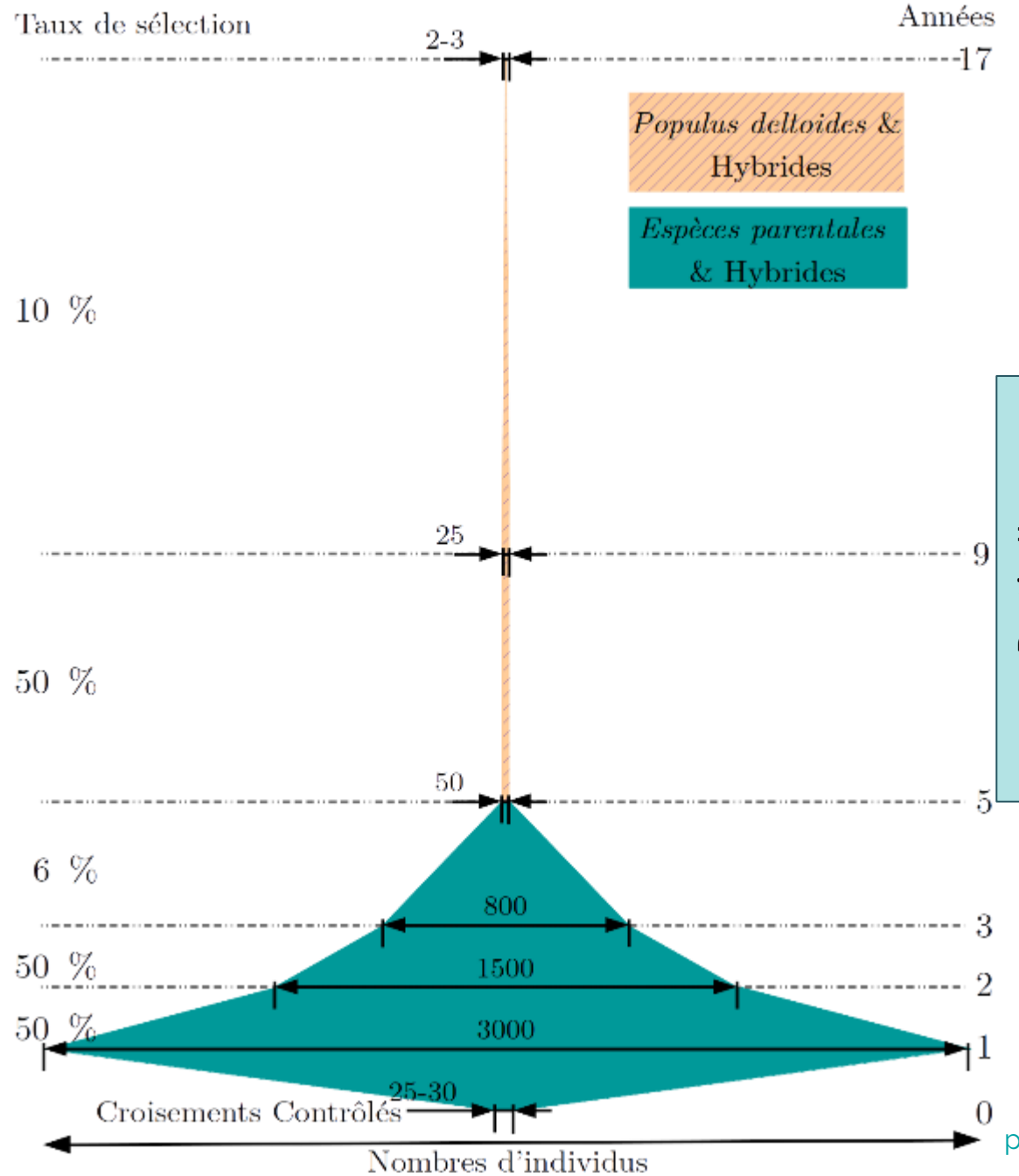
Populetum mono arbre :

- Résistance pathogènes / ravageurs
- Architecture

Pépinière :

- Croissance
- Phénologie
- angle de branche
- résistance à la rouille

Multiplication végétative



INRAE

M2 Parcours Plantes et Société

11 Septembre 2024 / Tours / Marie PEGARD

# ➤ Optimisation schéma de sélection

Qu'apporte la sélection génomique?

$$G = \frac{i r \sigma_A}{L}$$

Cycle Long



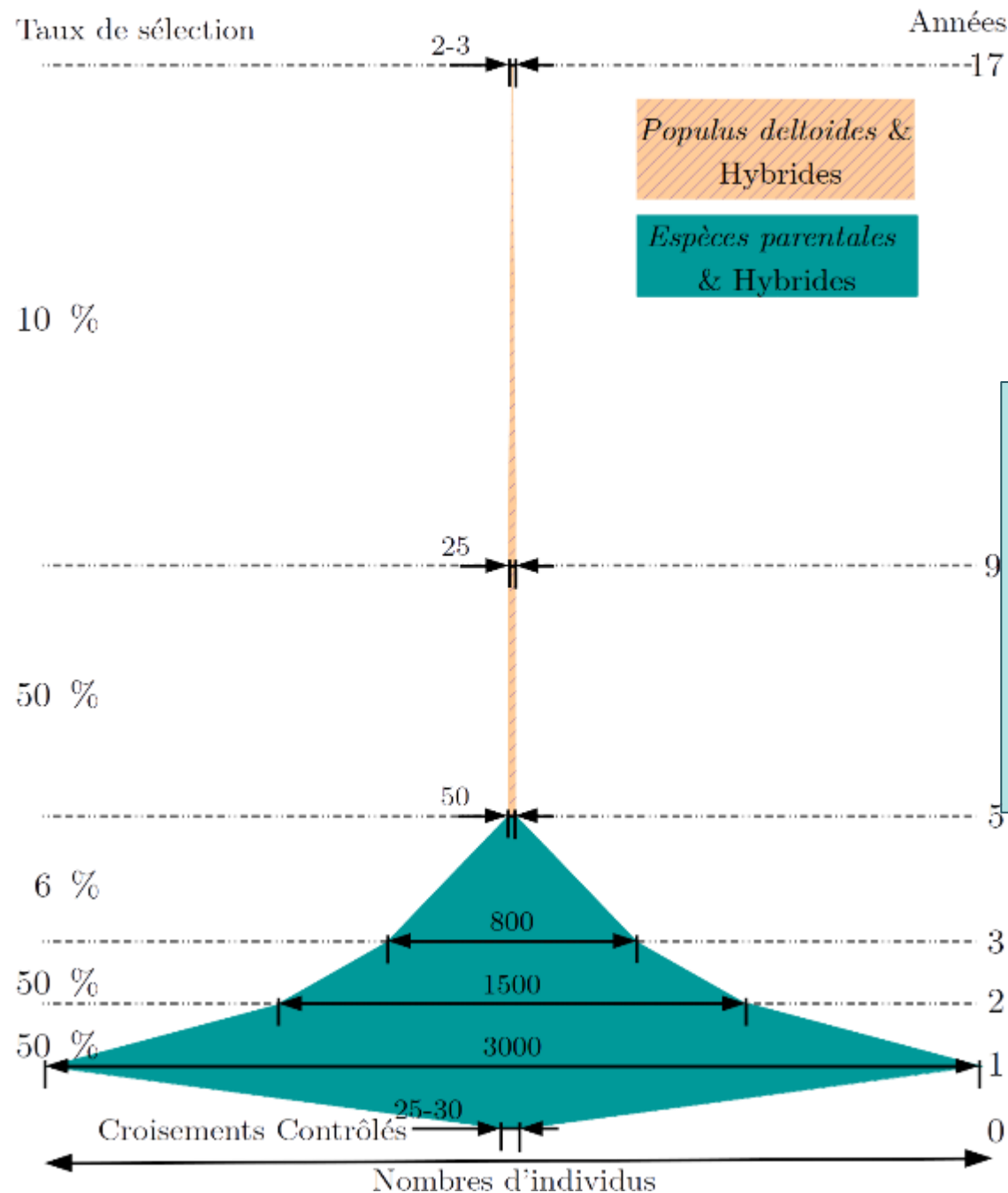
Forte intensité de sélection



Précision de sélection la plus faible

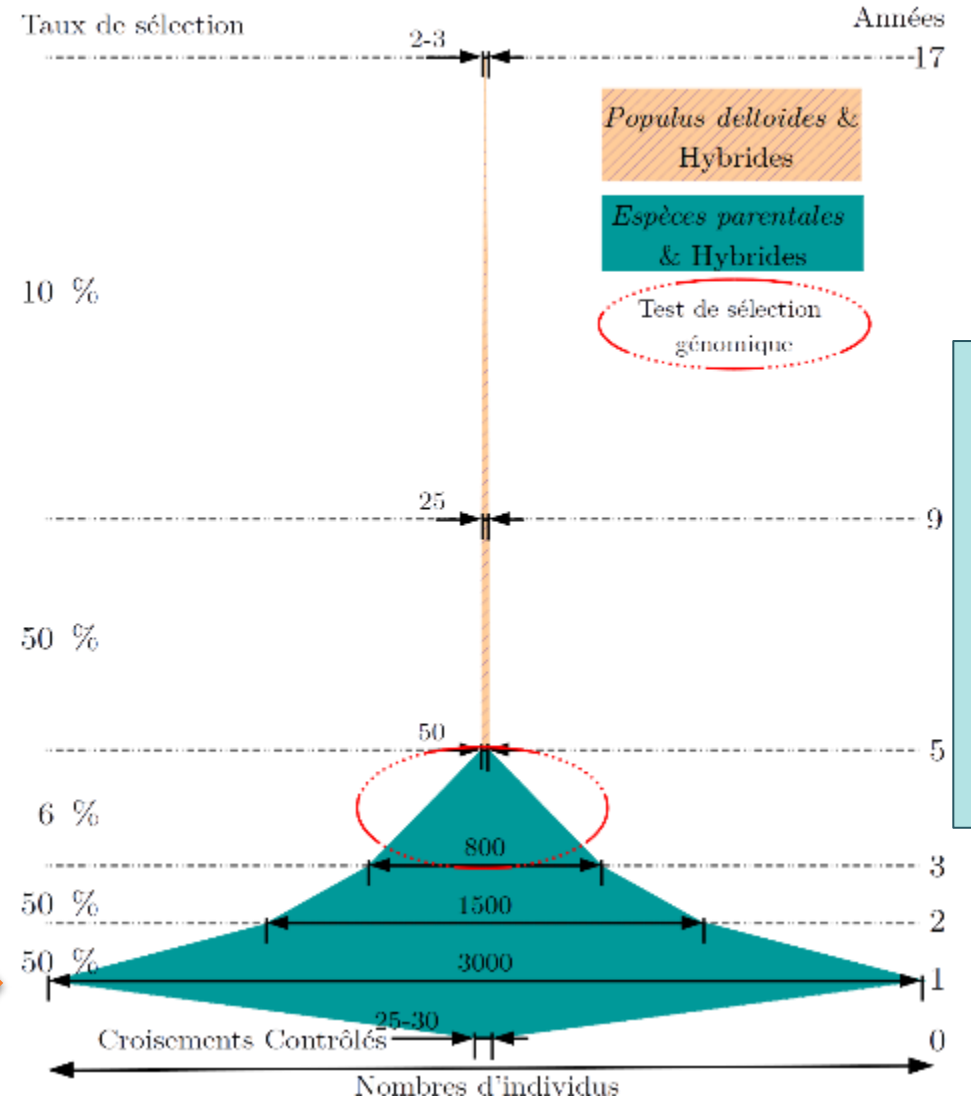


Variabilité génétique la plus forte



# ➤ Optimisation schéma de sélection

Comment intégrer la sélection génomique?

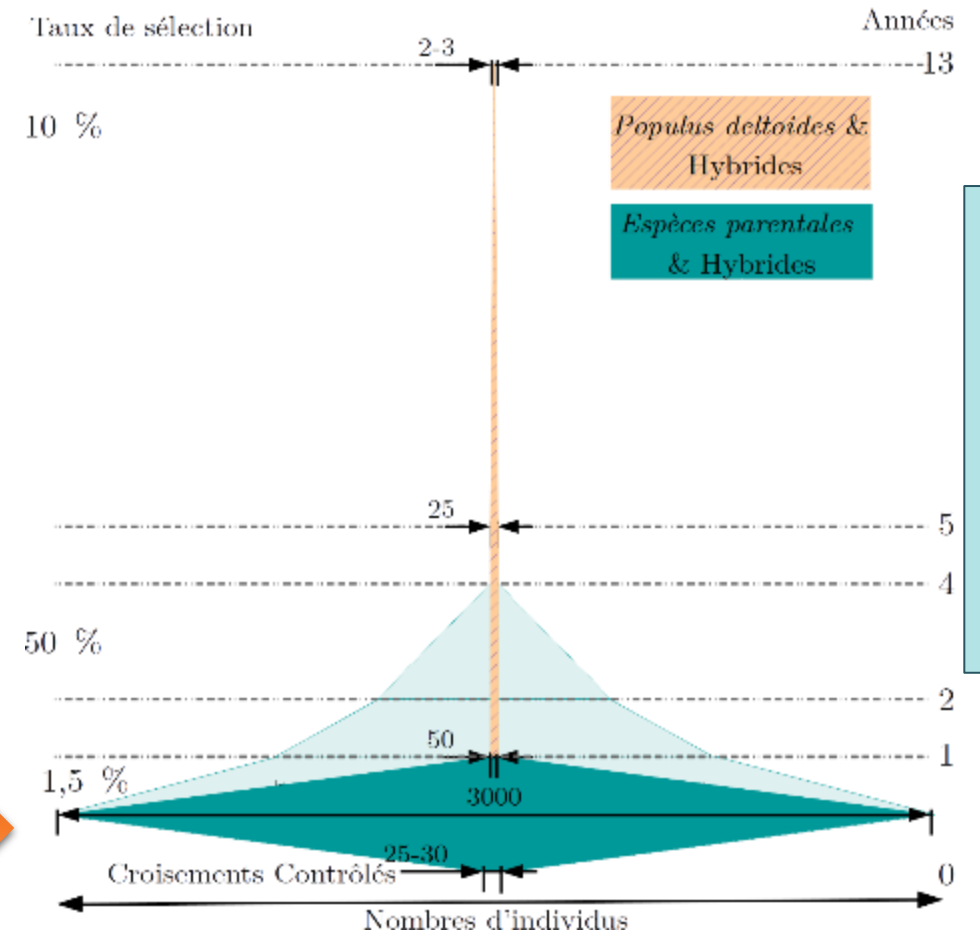


# ➤ Optimisation schéma de sélection

Comment intégrer la sélection génomique?

A court terme :

- Au stade plantule → grande variabilité
- Combiner une approche multi- caractères
- Améliorer la précision
- P. nigra*



# ➤ Optimisation schéma de sélection

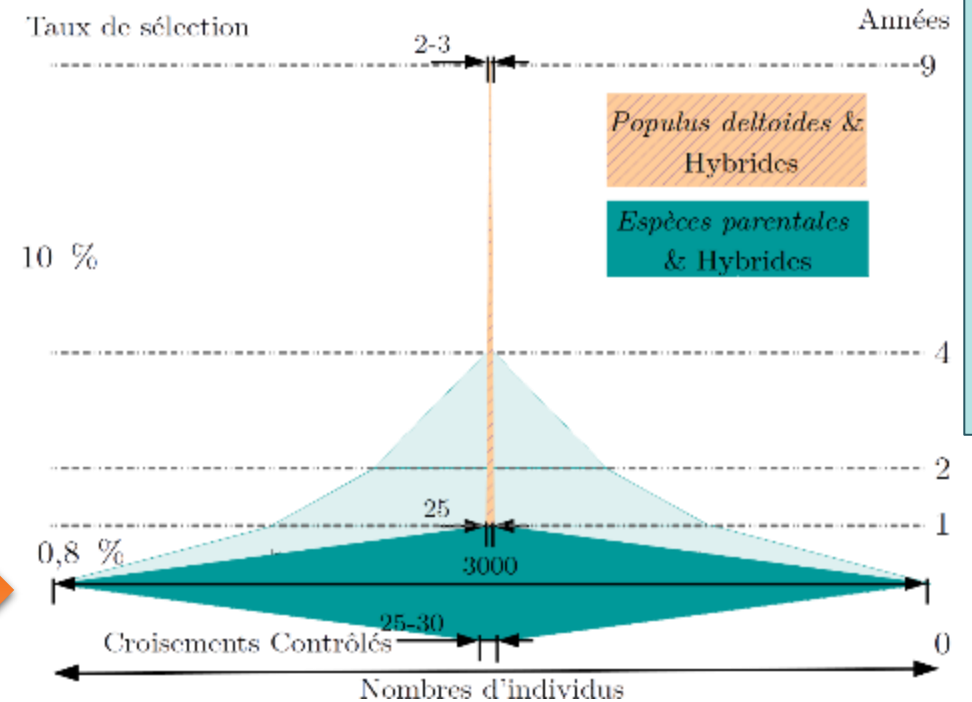
Comment intégrer la sélection génomique?

A court terme :

Au stade plantule → grande variabilité  
Combiner une approche multi- caractères  
Améliorer la précision  
*P. nigra*

A plus Long terme :

*P. deltoides*  
Hybrides  
Intégrer des caractères évalués plus  
tardivement ou plus coûteux





INRAE

➤ Gestion / Conservation ressource génétique

Partie 3

## ➤ Qu'est-ce qu'une ressource génétique

Selon le Ministère de l'enseignement supérieur, de la recherche et de l'innovation

Selon la Convention sur la Diversité Biologique (CDB) adoptée en 1992 lors du Sommet de Rio, les **ressources génétiques** sont un "**matériel** d'origine végétale, animale, microbienne ou autre, **contenant des unités fonctionnelles de l'hérédité**".

Ces matériels génétiques ont une valeur effective ou potentielle.

Les "connaissances traditionnelles associées aux ressources génétiques" sont détenues par une communauté d'habitants autochtone ou locale. Leur **utilisation** présente un **intérêt scientifique** ou **technique** précisée dans les **conditions convenues d'un commun accord** entre fournisseur et utilisateur.

Dans les années 90, les ressources génétiques sont devenues des **composants stratégiques** pour les industriels lorsqu'ils se sont lancés dans le développement des biotechnologies. Les industries qui utilisent le plus de ressources génétiques sont : les industries pharmaceutique, biotechnologique, cosmétique et agro-alimentaire.

**Gérer les ressources génétiques** consiste à inventorier, caractériser et évaluer, conserver, régénérer et diffuser.

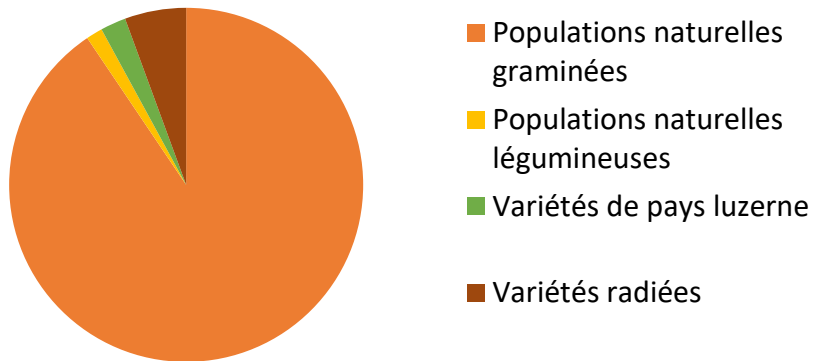


## ➤ Comment et par qui sont-elles gérées?

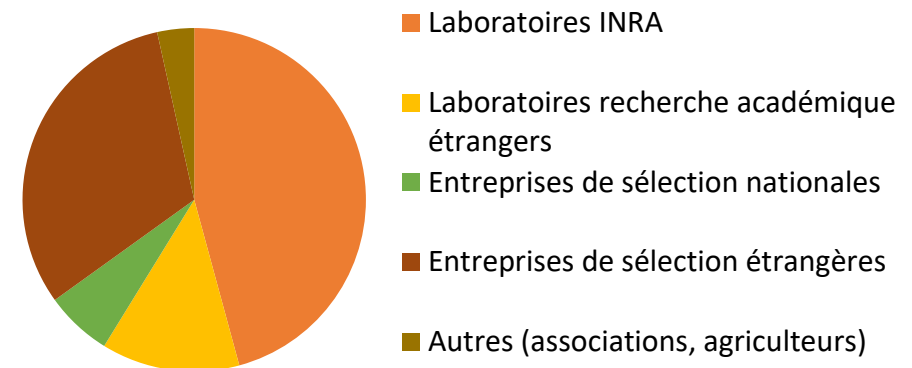
### Le CRB Prairies (Espèces Fourragères et à Gazon)

- **conserve** sous forme de graine, en chambre froide et au congélateur, **1589 accessions** de pop. naturelles de graminées (*Lolium*, *Festuca*, *Dactylis*) et légumineuses (*Medicago*, *Trifolium*, *Vicia*) et quelques variétés.
- maintient d'**une collection nationale** de **731 accessions** dans le cadre d'un **réseau national**
- représente le réseau national dans les activités du **réseau européen des ressources génétiques ECP GR** (groupe de travail espèces fourragères) et contribue à l'établissement de **la collection européenne AEGIS** pour les espèces fourragères.

#### Composition de la Collection nationale



#### Env. 600 échantillons ont été distribués de 2009 à 2019






# ➤ Comment et par qui sont-elles gérées?

## RÉGLEMENTATION APA

La réglementation APA encadre l'accès et l'utilisation des **R**essources **G**énétiques (RG) et des **C**onnaissances **T**raditionnelles qui leurs sont **A**ssociées (CTA).

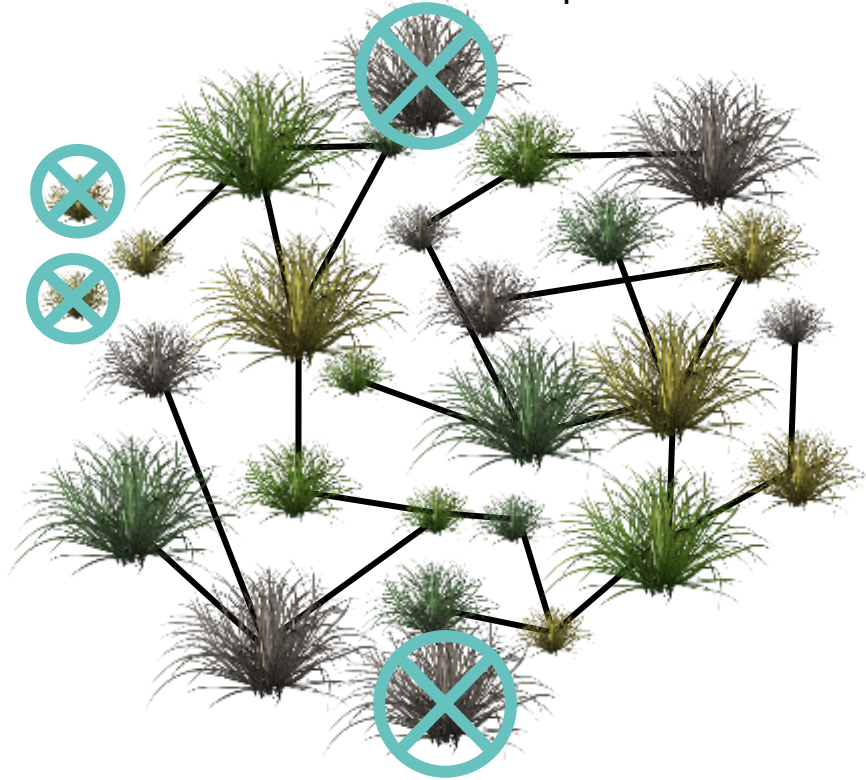
 Origine : elle découle de la **Convention sur la Diversité Biologique (Rio, 1992)** adoptée sous l'égide des Nations unies et son **Protocole additionnel (Nagoya, 2010)**.

Ces textes poursuivent **trois grands objectifs** :

-  Éviter l'épuisement des ressources biologiques
-  Assurer leur utilisation durable
-  Garantir entre un partage juste et équitables des avantages (APA) découlant de l'utilisation des ressources génétiques.

## ➤ Vérification d'identité ou duplicatas

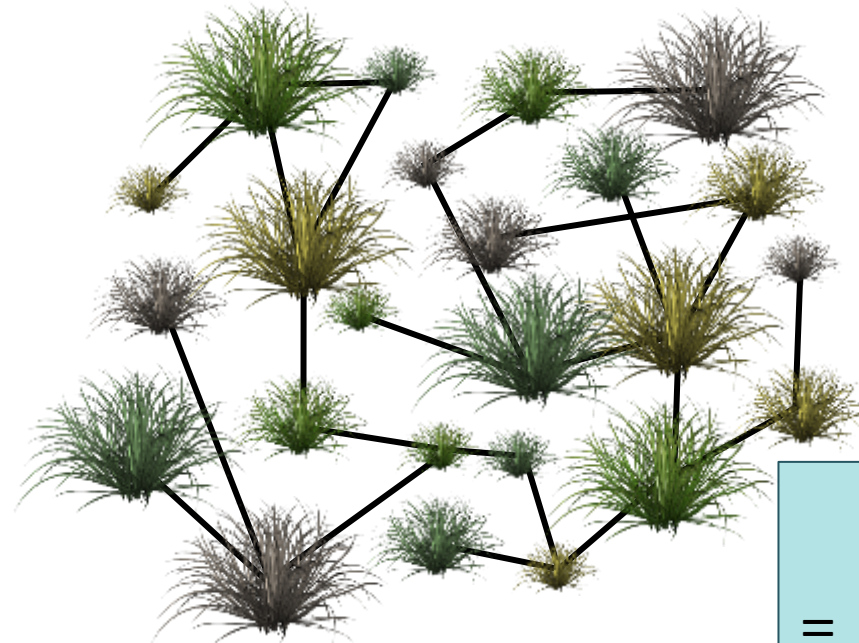
Collection complète



Génotypage de la collection

- SSR
- SNP
- GBS

Suppression des duplicatas ou ils sont rassemblés sous le même identifiant



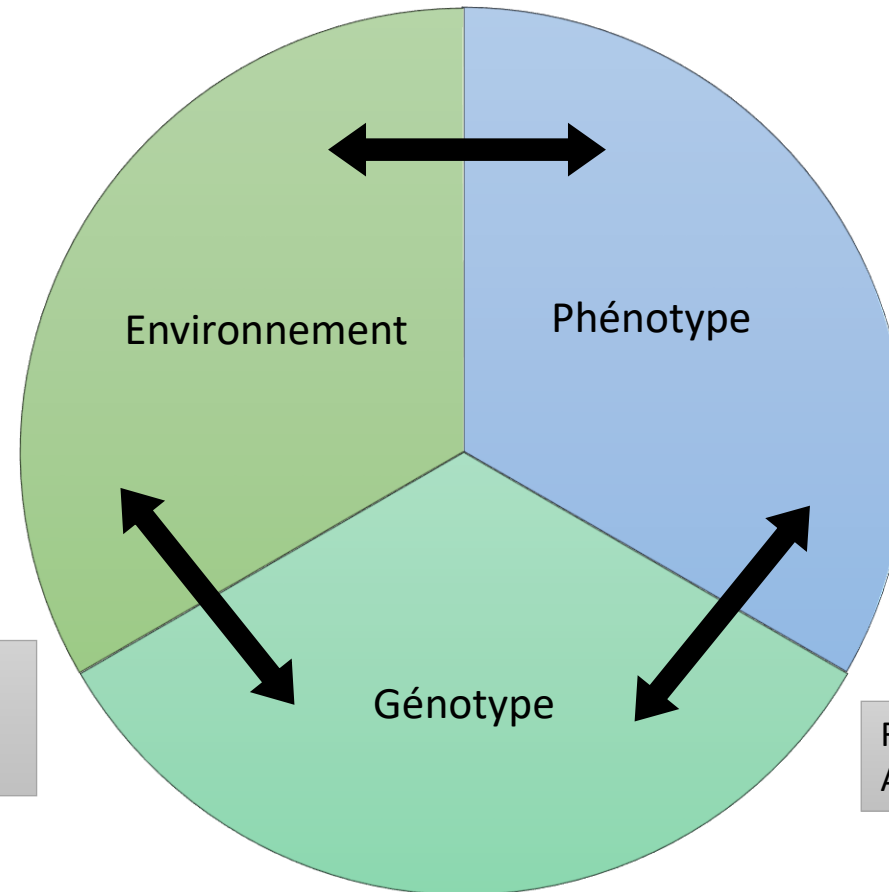
# ➤ Caractérisation des ressources génétiques

Phénotypique, moléculaire ....

Documentation de l'origine des populations

Données climatiques

Recherche Genome Environnement Association (GEA) et signature de la sélection



Phénotypage peu coûteux sur une grande collection

Phénotypage très coûteux sur seulement une partie de la collection

Phénotypage haut débit

- NIRS
- Drone

Recherche QTL-GWAS (Genome Wide Association Study)

Génotypage :

- GBS
- Capture

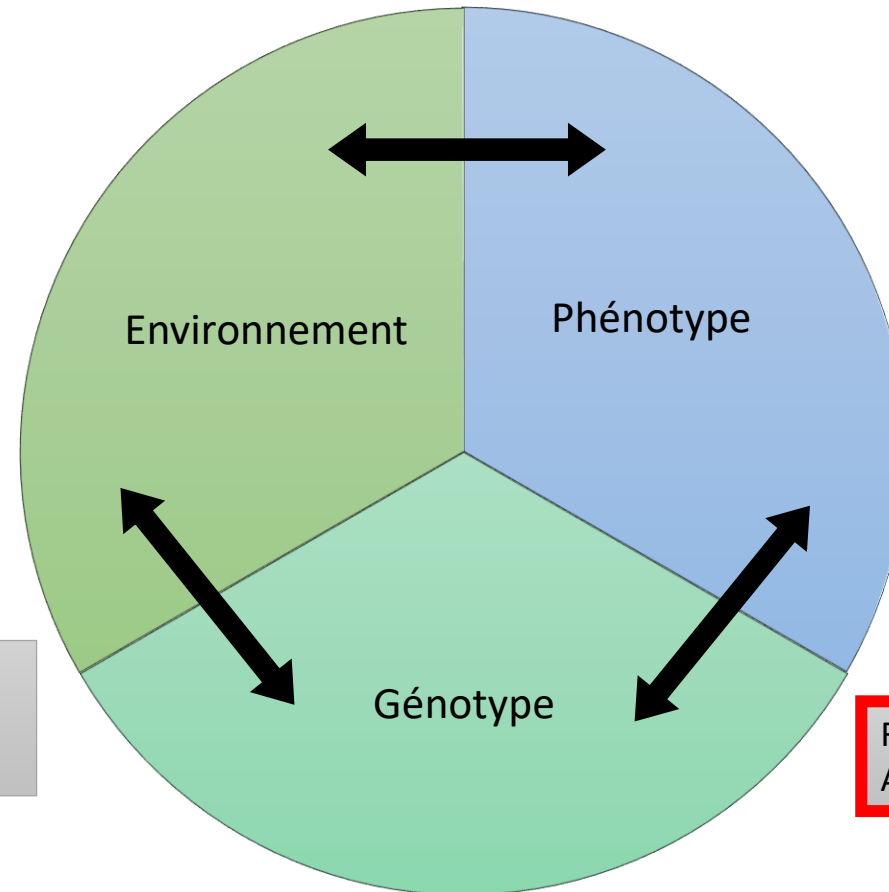
# ➤ Caractérisation des ressources génétiques

Phénotypique, moléculaire ....

Documentation de l'origine des populations

Données climatiques

Recherche Genome Environnement Association (GEA) et signature de la sélection



Phénotypage peu coûteux sur une grande collection

Phénotypage très coûteux sur seulement une partie de la collection

Phénotypage haut débit

- NIRS
- Drone

Recherche QTL-GWAS (Genome Wide Association Study)

Génotypage :

- GBS
- Capture



INRAE

M2 Parcours Plantes et Société

11 Septembre 2024 / Tours / Marie PEGARD

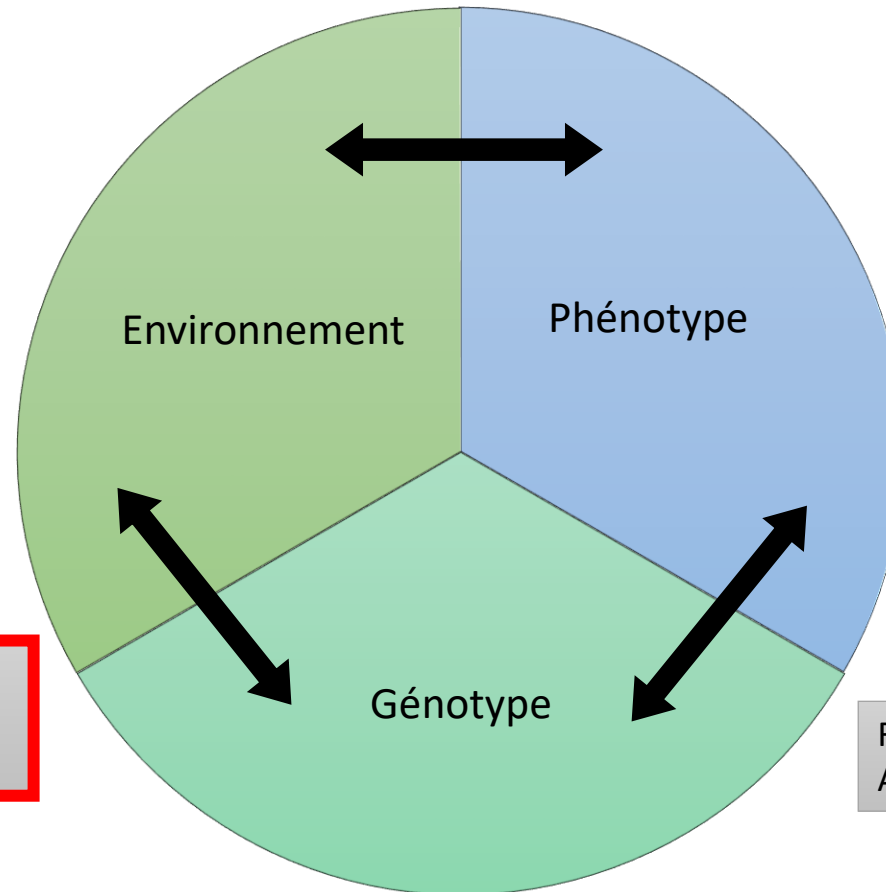
# ➤ Caractérisation des ressources génétiques

Phénotypique, moléculaire ....

Documentation de l'origine des populations

Données climatiques

Recherche Genome Environnement Association (GEA) et signature de la sélection



Phénotypage peu coûteux sur une grande collection

Phénotypage très coûteux sur seulement une partie de la collection

Phénotypage haut débit

- NIRS
- Drone

Recherche QTL-GWAS (Genome Wide Association Study)

Génotypage :

- GBS
- Capture



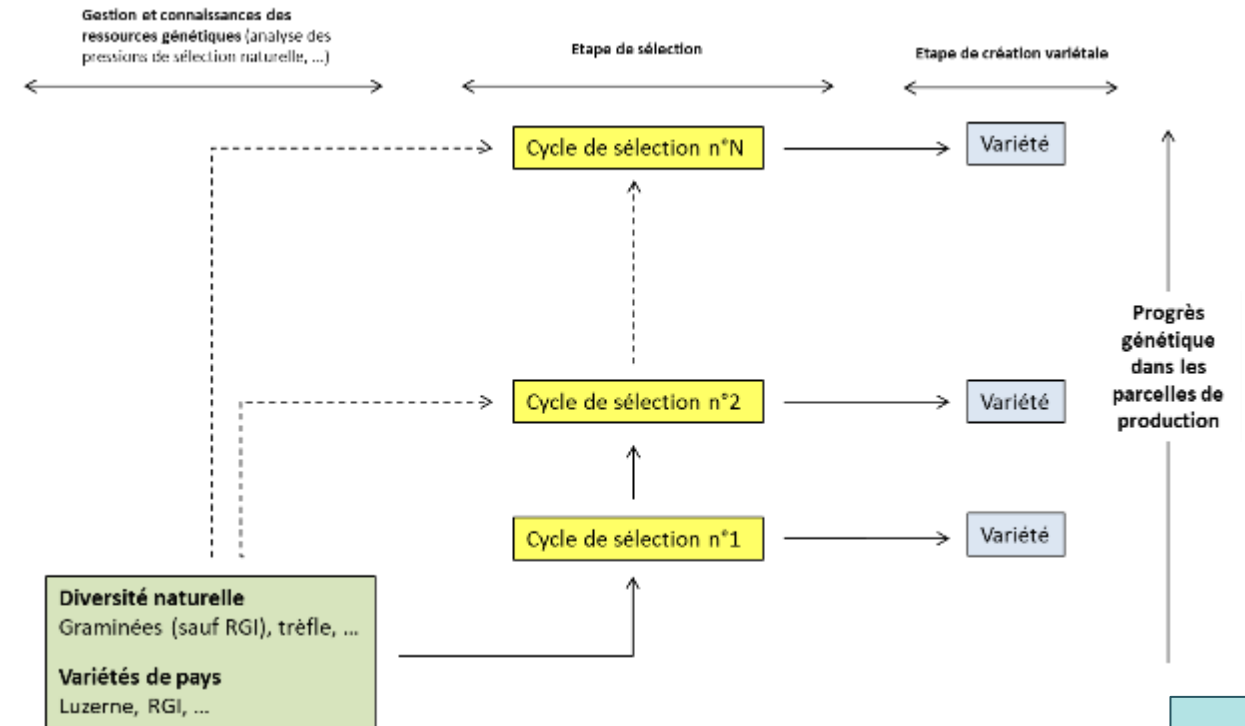
# ➤ Caractérisation des ressources génétiques

Pourquoi rechercher la diversité moléculaire adaptative dans les populations naturelles ?

La diversité adaptative est le résultat de pressions de sélection naturelle qui tendent à adapter les phénotypes naturels aux conditions environnementales locales

Il est de plus en plus évident que la majeure partie de la diversité phénotypique naturelle est adaptative (Rieseberg et al., 2002 ; Rieseberg et al., 2004).

La diversité phénotypique naturelle constitue le matériel de base pour la sélection végétale (ressources génétiques).



- Rechercher la diversité moléculaire adaptative, c'est rechercher la diversité moléculaire potentiellement utilisable pour la sélection végétale.
- La détection de la diversité adaptative directement au niveau moléculaire peut révéler la fonction adaptative de caractères qui ne sont généralement pas évalués par les sélectionneurs (et qui présentent un intérêt potentiel pour la sélection).

## ➤ Caractérisation de nouvelles ressources génétiques

Détection de la diversité adaptative dans les polymorphismes du génome : contexte

La **démographie** et l'**histoire des populations** impriment des modèles structurels similaires à **tous les loci** dans les génomes par le biais de la **dérive** et de la **migration**.

La **sélection** n'affecte qu'un **nombre limité** de loci

Les loci neutres situés à proximité des loci sélectionnés tendent à présenter des schémas spécifiques de variation allélique (zones de **balayage sélectif**).

Loci soumis à des pressions de sélection adaptative peuvent être détectés par la moyenne des marqueurs neutres présentant :

- un spectre de fréquence allélique typique des zones de balayage sélectif : méthodes de **génétique des populations**
- une distribution allélique corrélée aux variations environnementales : **statistiques d'association** (QTL, GWAS ...)



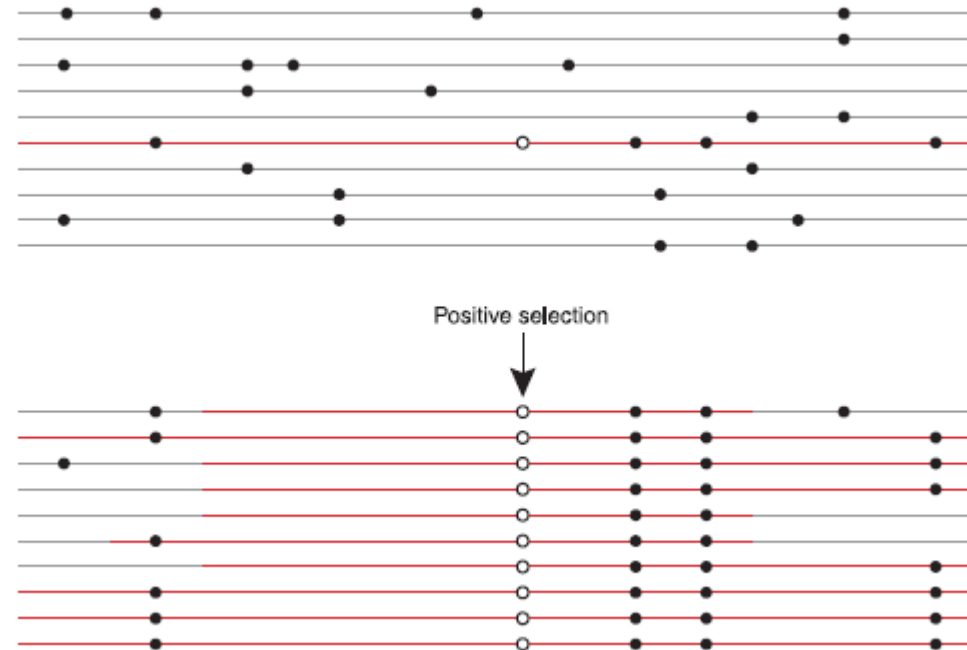
## ➤ Rappel sur le balayage sélectif

Genetic hitch-hiking along a recombining chromosome in a natural population

(from Storz, 2005)

### Diversité moléculaire au voisinage d'un locus sélectionné

Une mutation favorable (présentant un avantage sur le plan de la fitness) apparaît et est rapidement amenée à se fixer (sélection positive).



Les mutations neutres initialement en déséquilibre de liaison avec la mutation favorable subissent une augmentation spectaculaire de leur fréquence.

- Lignes horizontales : un échantillon de population de chromosomes homologues
- Cercle ouvert : mutation favorable
- Cercles remplis : mutations neutres

➔ Balayage sélectif : une fraction importante de l'haplotype ancestral s'est fixée.

# ➤ Etude de la structuration génétique

## Mesure de la différenciation entre (sous) populations: le $F_{ST}$

Une population divisée en plusieurs sous-populations:

→ les sous-populations sont génétiquement différenciées si elles présentent des fréquences alléliques différentes (dérive, sélection, isolement génétique entre sous-populations)

→ le taux d'hétérozygotes dans la population globale est inférieur à celui qu'on aurait en situation de panmixie

$$F_{ST} = (H_s - H_I) / H_s$$

$H_s$  = hétérozygotie attendue dans la population globale à l'équilibre Hardy-Weinberg

$H_I$  = hétérozygotie attendue dans les sous-populations à l'équilibre Hardy-Weinberg

$F_{ST}$  entre 0 et 0,05 → différenciation faible

$F_{ST}$  entre 0,05 et 0,15 → différenciation moyenne

$F_{ST}$  entre 0,15 et 0,25 → différenciation importante

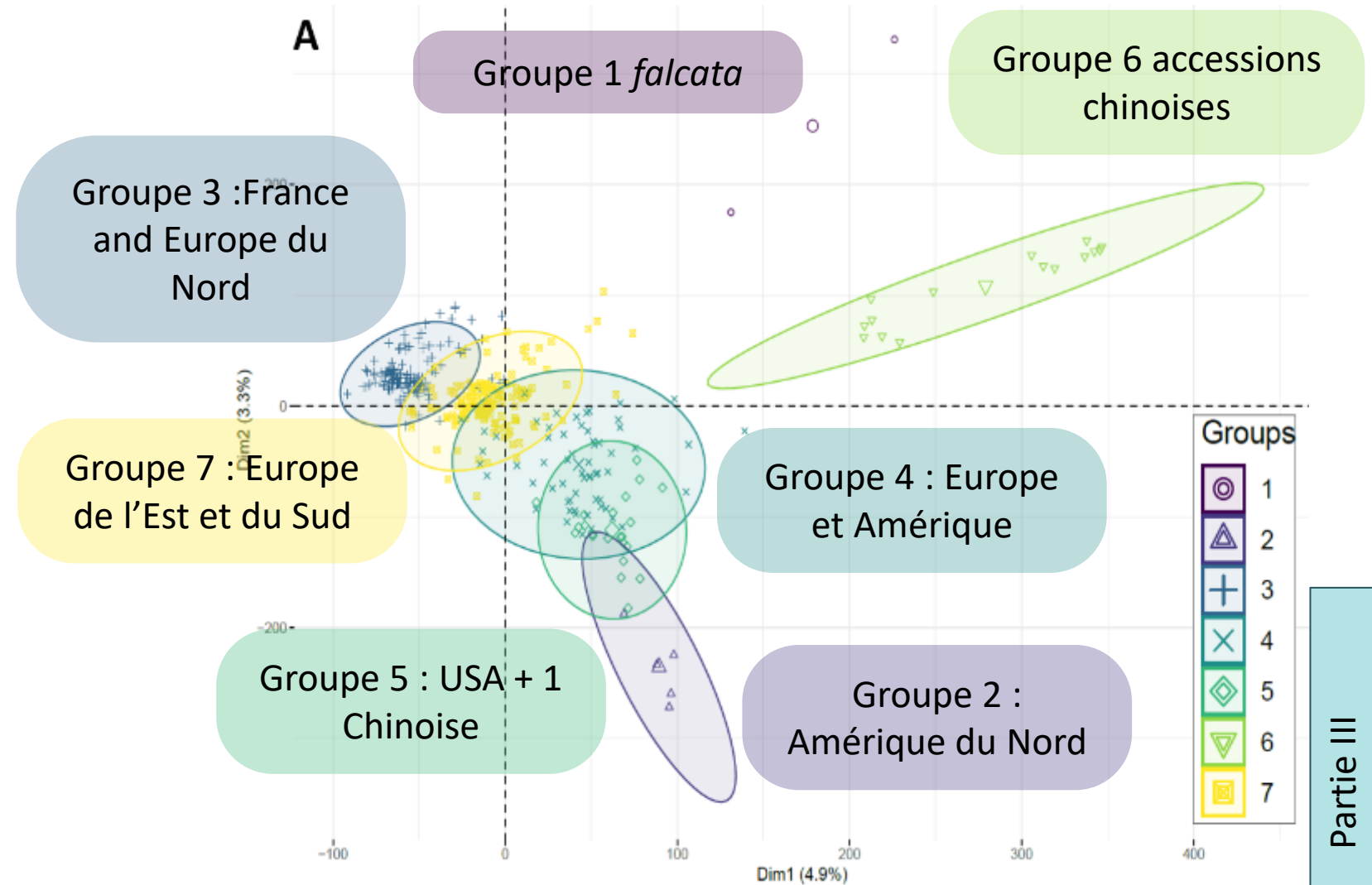
$F_{ST}$  supérieur à 0,25 → différenciation très forte



# ➤ Etude de la structuration génétique

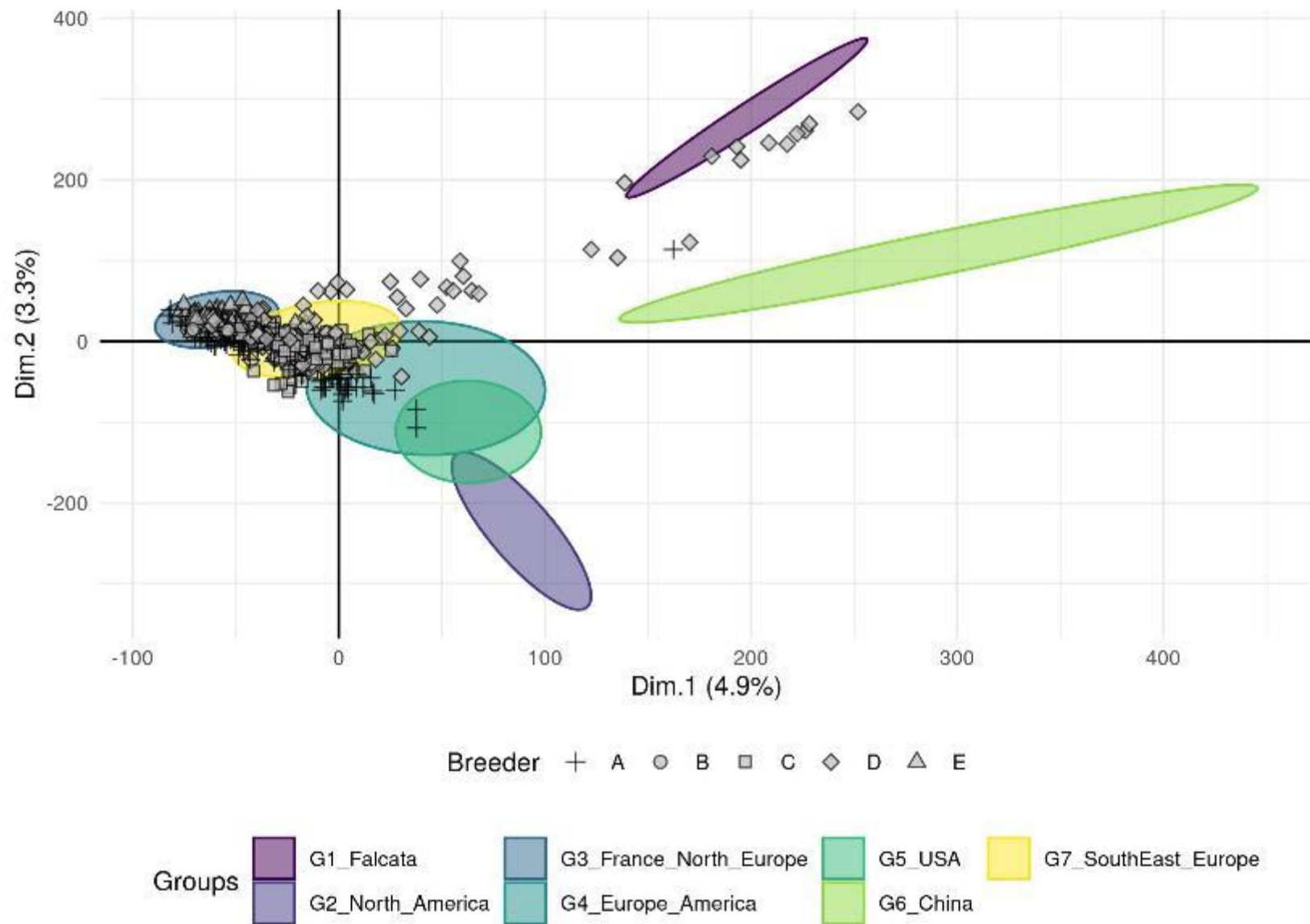
Exemple chez la luzerne

- Deux groups se distinguent clairement des autres
- Les 5 autres groups forment un continuum génétique
- Les accessions européennes et américaines se séparent entre plusieurs groups



# ➤ Etude de la structuration génétique

Exemple chez la luzerne



INRAE

M2 Parcours Plantes et Société

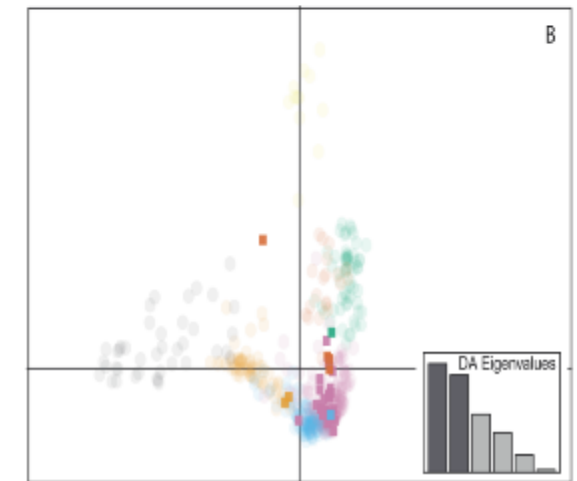
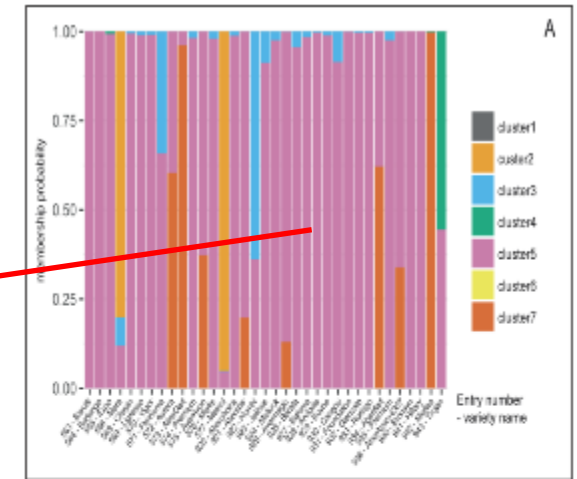
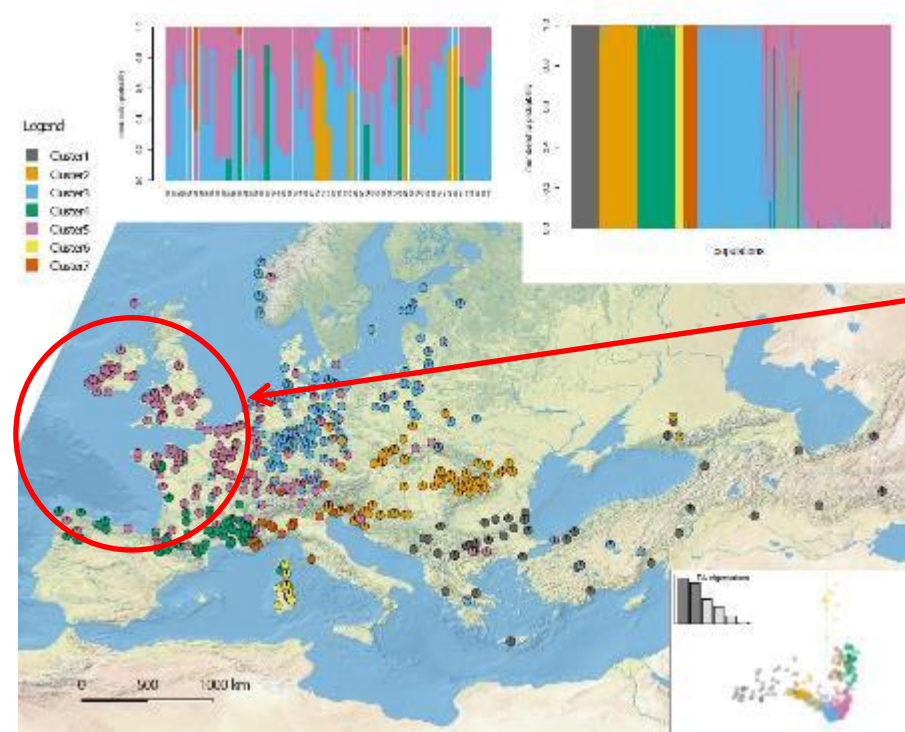
11 Septembre 2024 / Tours / Marie PEGARD

# ➤ Etude de la structuration génétique

Exemple chez le ray-grass anglais

## Résultats

- Les programmes de sélection n'ont utilisé qu'une partie limitée de la diversité naturelle de *L. perenne* (groupe de l'Europe de l'Ouest).
- Les cultivars sélectionnés ont tendance à remplacer la diversité naturelle dans toute l'Europe.
- Risque de perte de la diversité naturelle (adaptative)



### *Additional genotyping of L. perenne cultivars:*

- 32 cultivars representing the diversity of cvs bred in Europe and New-Zealand
- Cultivars projected as additional individuals on the DAPC of natural populations



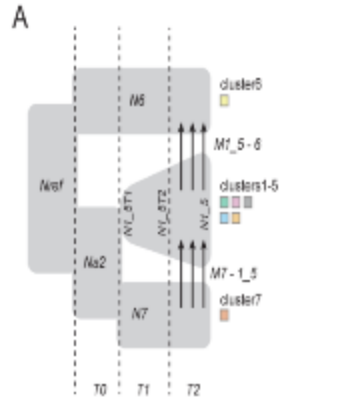
INRAE

M2 Parcours Plantes et Société

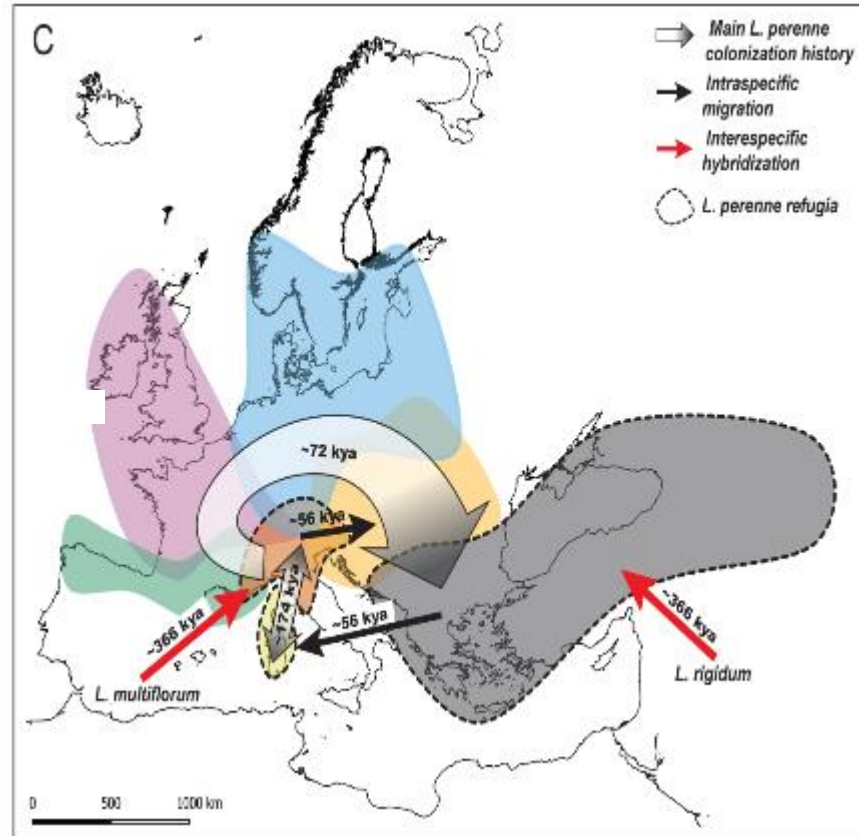
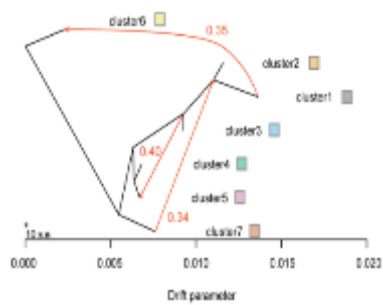
11 Septembre 2024 / Tours / Marie PEGARD

# ➤ Reconstruction phylogéographique

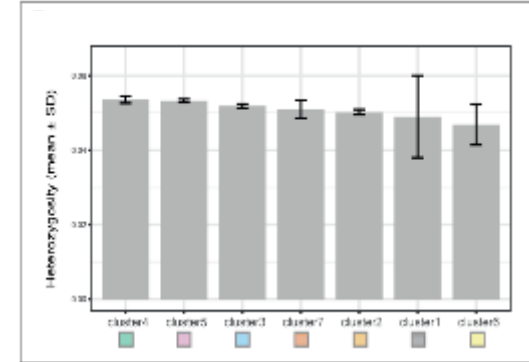
Exemple chez le ray-grass anglais



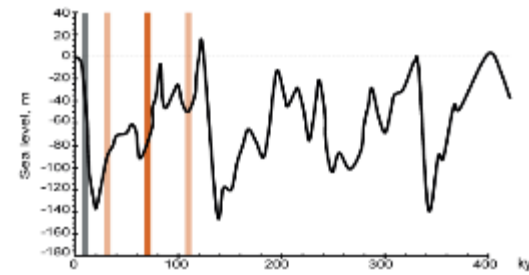
*δαδι* Gutenkunst, et al. (2009)



*Lolium perenne* phylogeography and expansion inferred from 507 583 nuclear SNPs



Heterozygosity within populations



Würm glaciation 12 – 110 kya, from Rabineau et al. (2006)

- L'expansion de l'espèce antérieure à l'apparition de l'agriculture en Europe
- La structure spatiale actuelle a été façonnée par les variations climatiques tout au long de l'ère quaternaire.





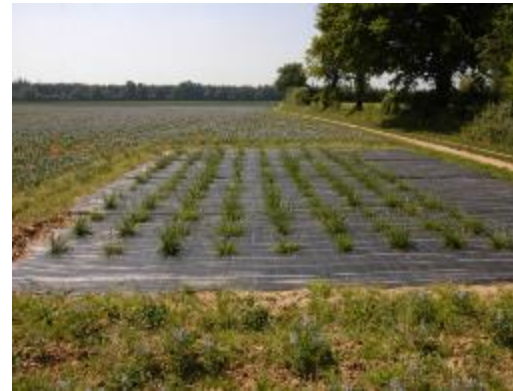
# ➤ Détection allèles adaptatifs

Exemple chez le ray-grass anglais projet Grasslandscape

Objectif principal : Rapprocher la génomique du paysage et la génétique quantitative pour une adaptation régionale des prairies européennes au changement climatique.



*In situ sampling*

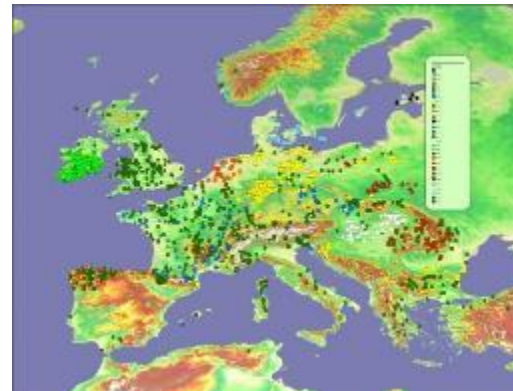


*Seed regeneration*



*Conservation seeds in cold*

*Opportunity: Large collections of natural populations of perennial ryegrass collected over the past decades and stored in genetic resource centers in European countries*



*About 3500 natural populations of perennial ryegrass conserved by European genetic resource centers*

# ➤ GrassLandscape

## Datasets

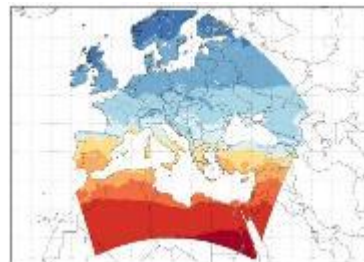


*Phenotyping*



*GBS genotyping*

469 accessions ➡



*Climatic variables*



### Three common garden with different climatic conditions : IPK, ILVO, INRAE

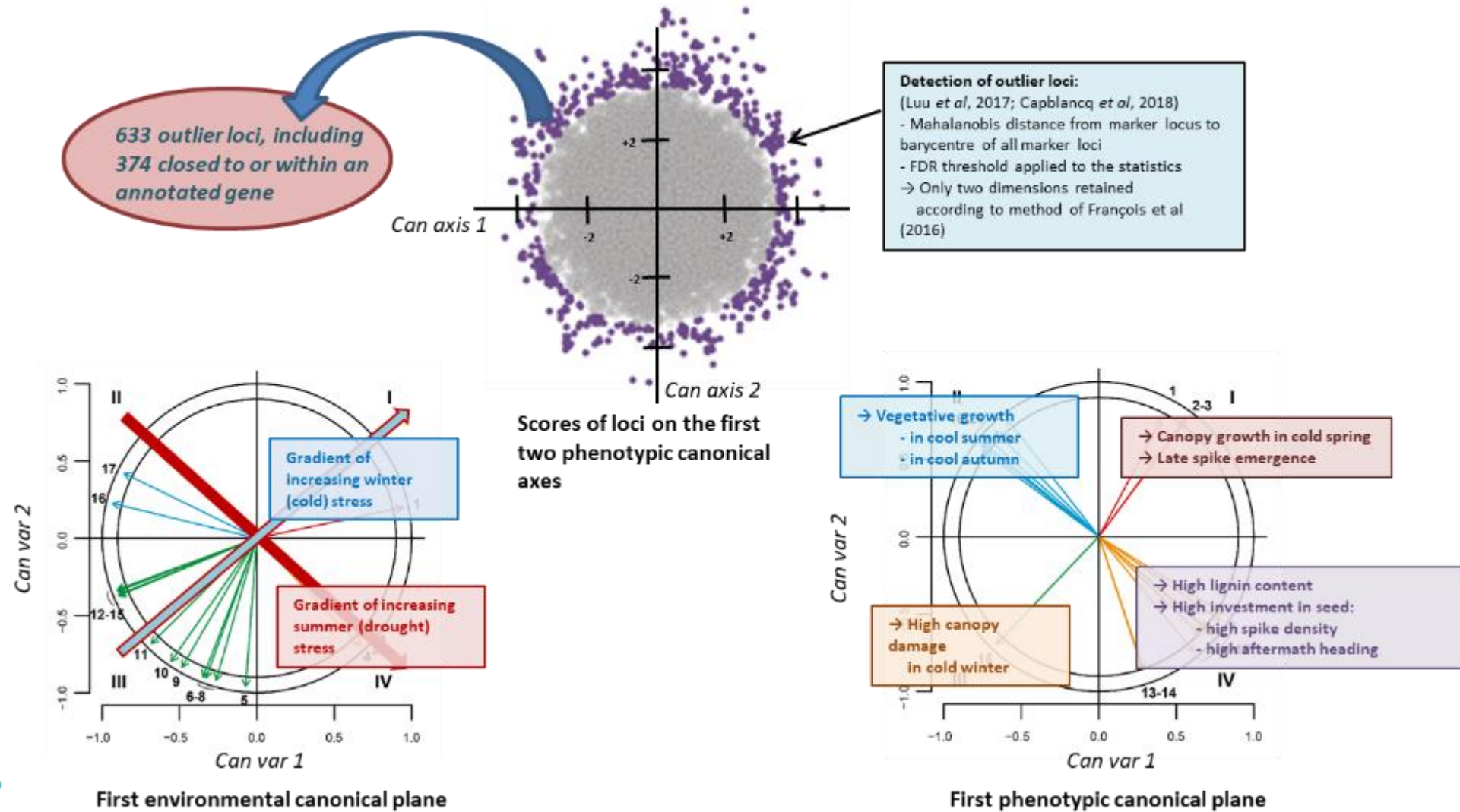
🕒 *Sowing on 04/2015 – Phenotyping up to 11/2018*

- 420 natural populations of perennial ryegrass and its relatives
- 32 selected perennial ryegrass varieties
- 3 complete blocks per location

- Growth dynamics
- Phenology, morphology, disease resistance
- Biochemical composition, 13C discrimination

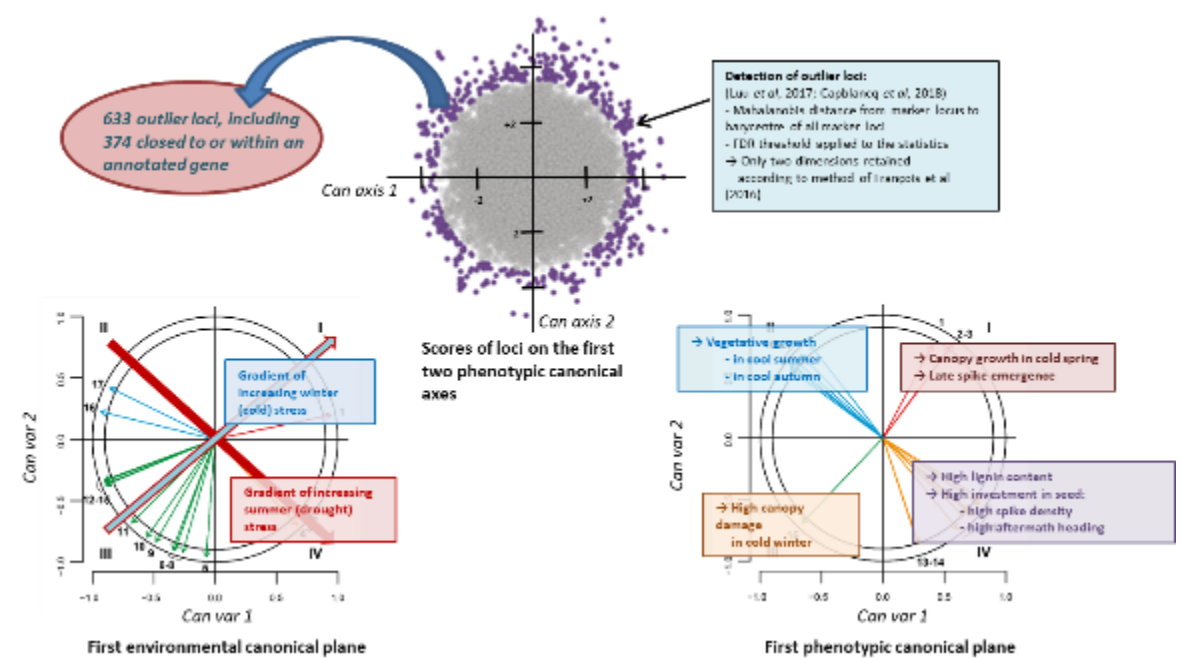
# ➤ Détection allèles adaptatifs

José Blanco-Pastor et al., 2021



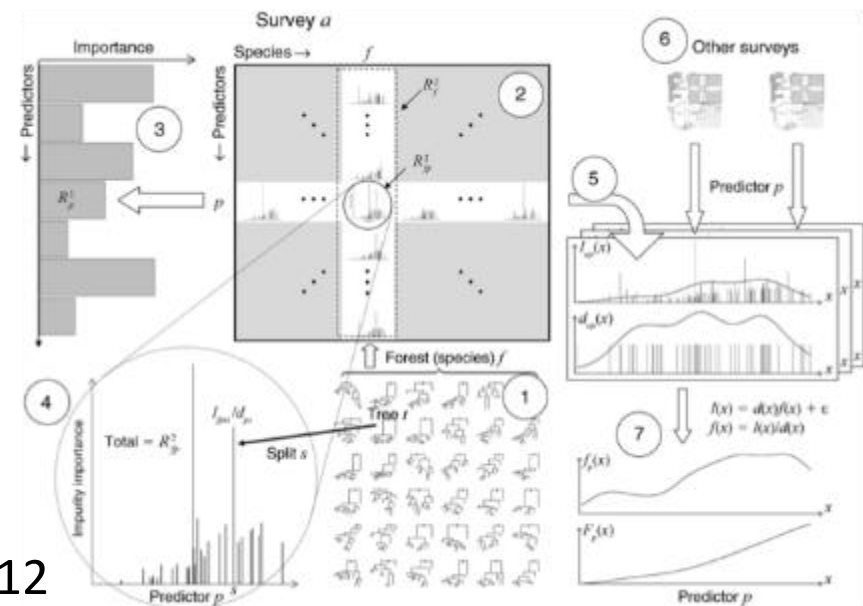
# ➤ Détection allèles adaptatifs

Autres méthodes



Méthode de détection de loci adaptatifs :

- Analyse CANCOR (Blanco-Pastor et al., 2019)
  - Détection allèles adaptatifs
  - Relations linéaires
- Gradient Forest et GDM (Fitzpatrick et al., 2015)
  - Prise en compte de la distance géographique
  - Relation linéaires et/ou non linéaires
  - Décrit et prédit l'évolution de la fréquence de l'allèle adaptatif



Ellis et al. 2012

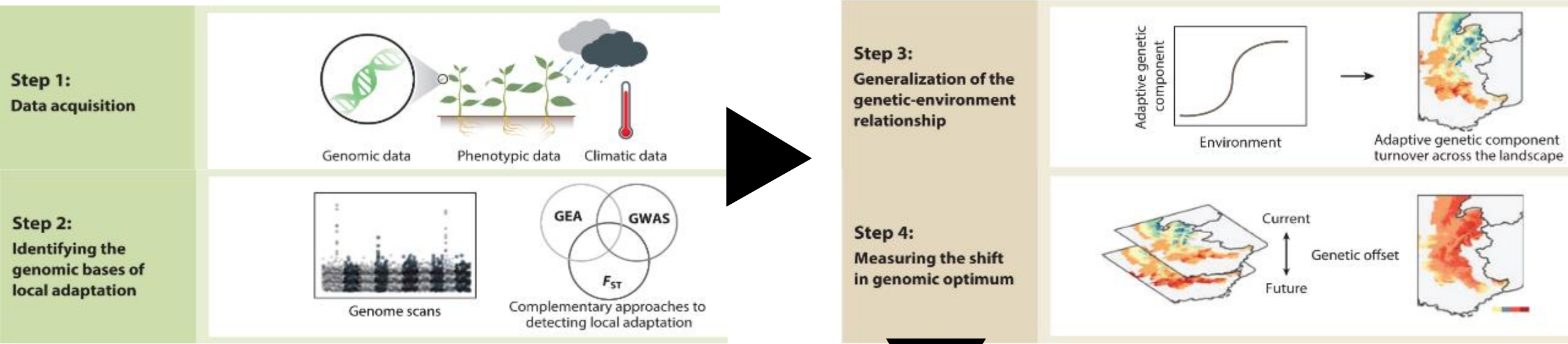


INRAE

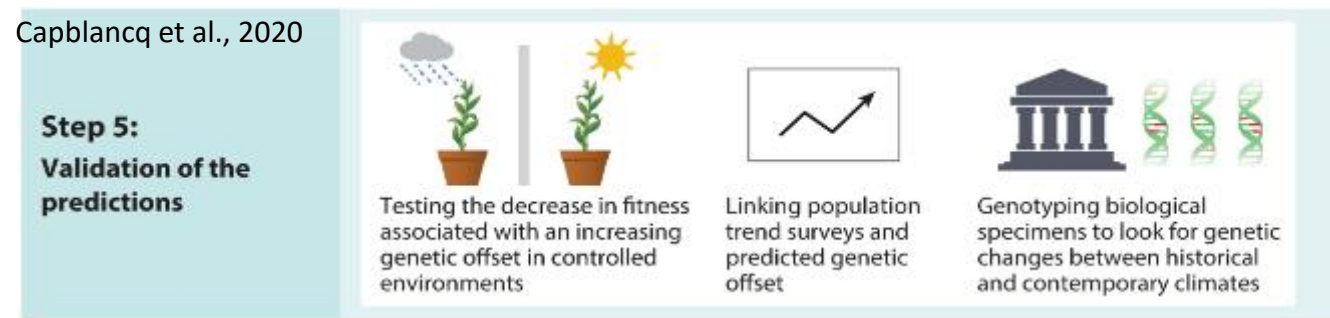
M2 Parcours Plantes et Société  
 11 Septembre 2024 / Tours / Marie PEGARD

# ➤ Évaluation de la vulnérabilité au changement climatique

## Genomic Offset : Methods



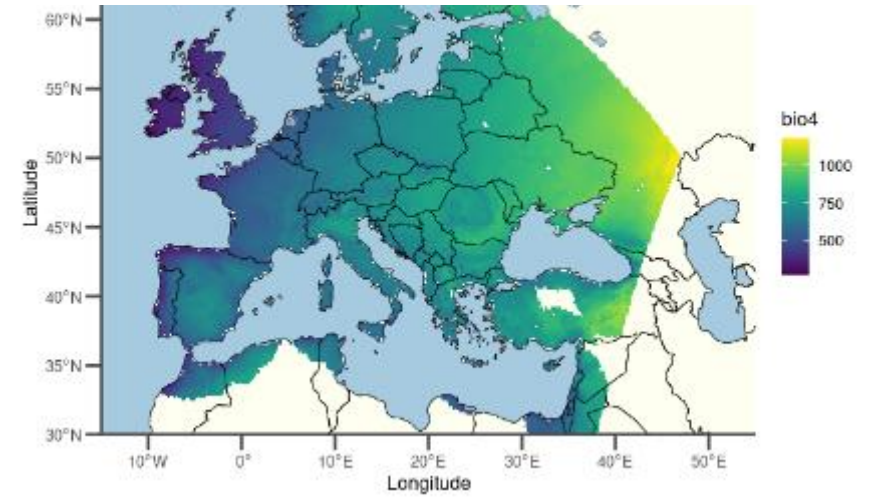
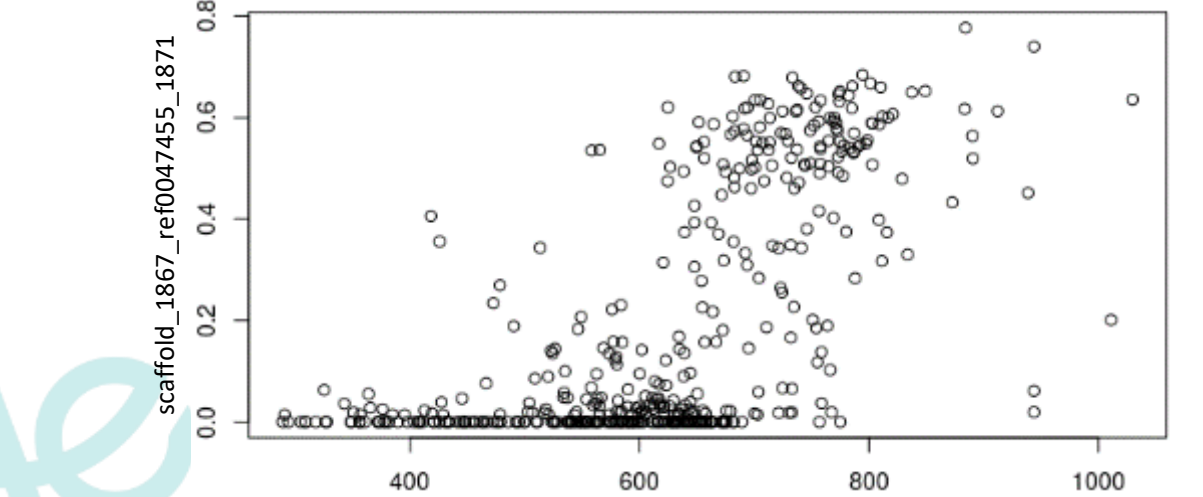
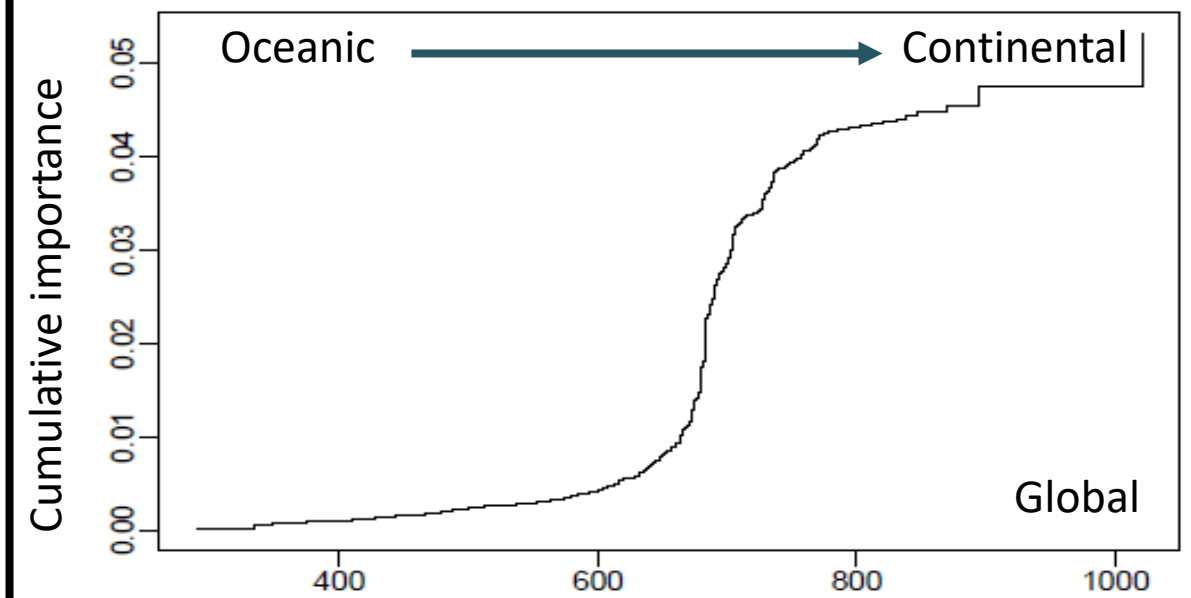
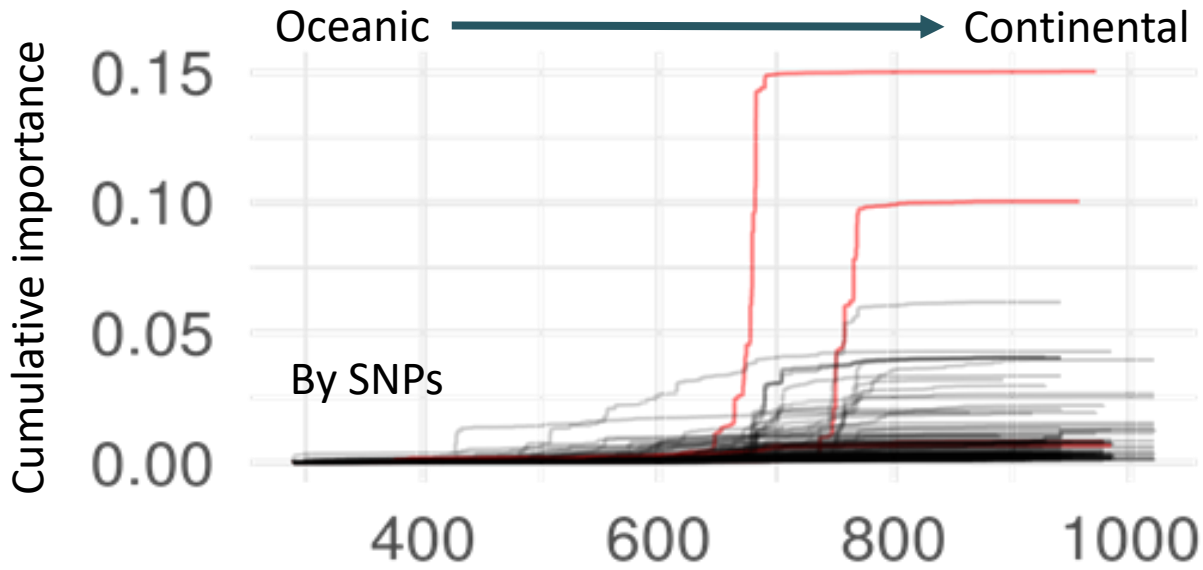
Capblancq et al., 2020



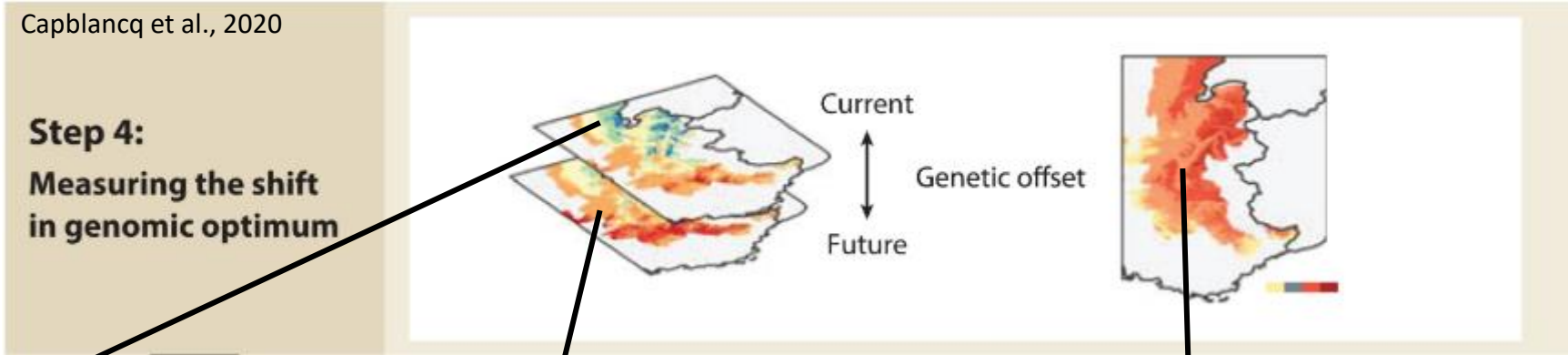
INRAE

# ➤ Genetic offset estimation on perennial ryegrass

Bio4 = Temperature seasonality (standard deviation of daily mean temperature)

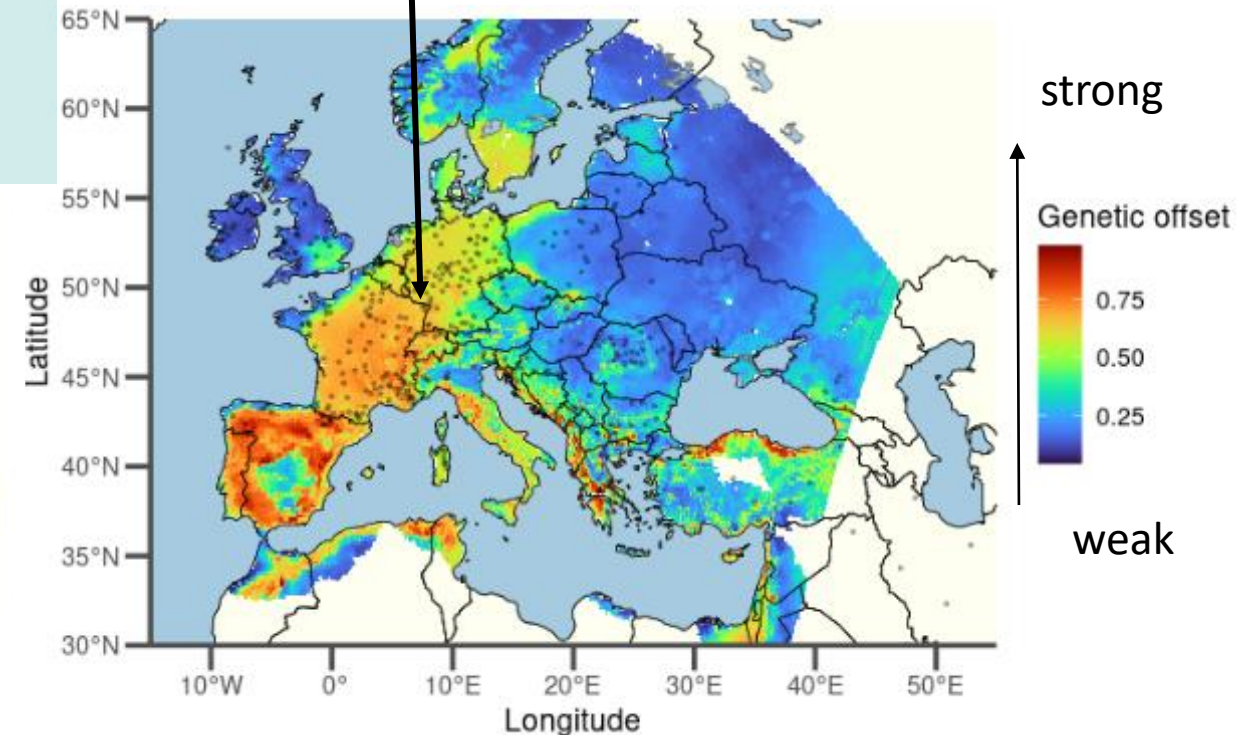
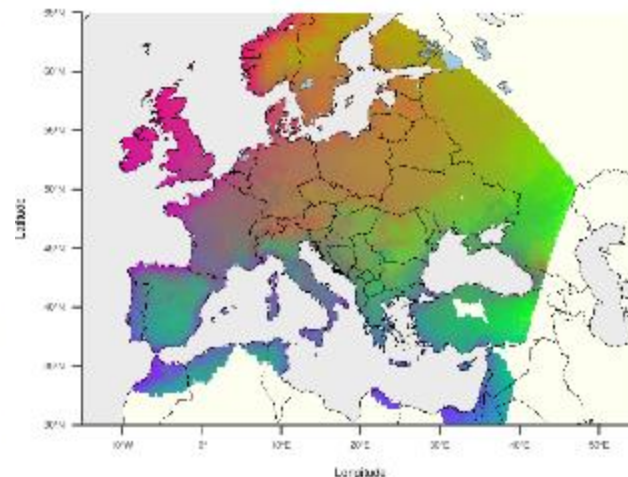
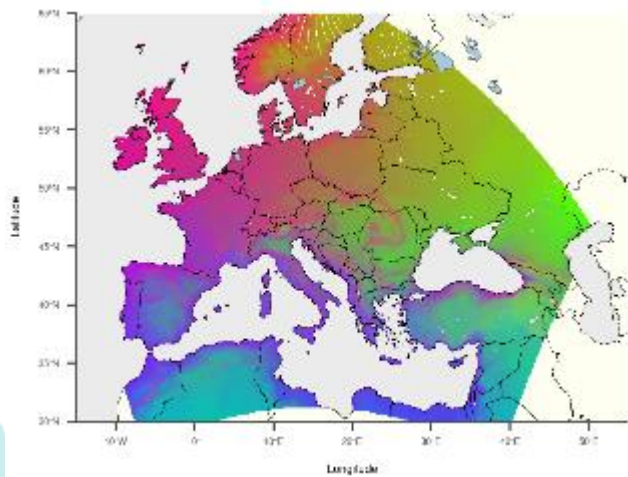


# ➤ Genetic offset estimation on perennial ryegrass



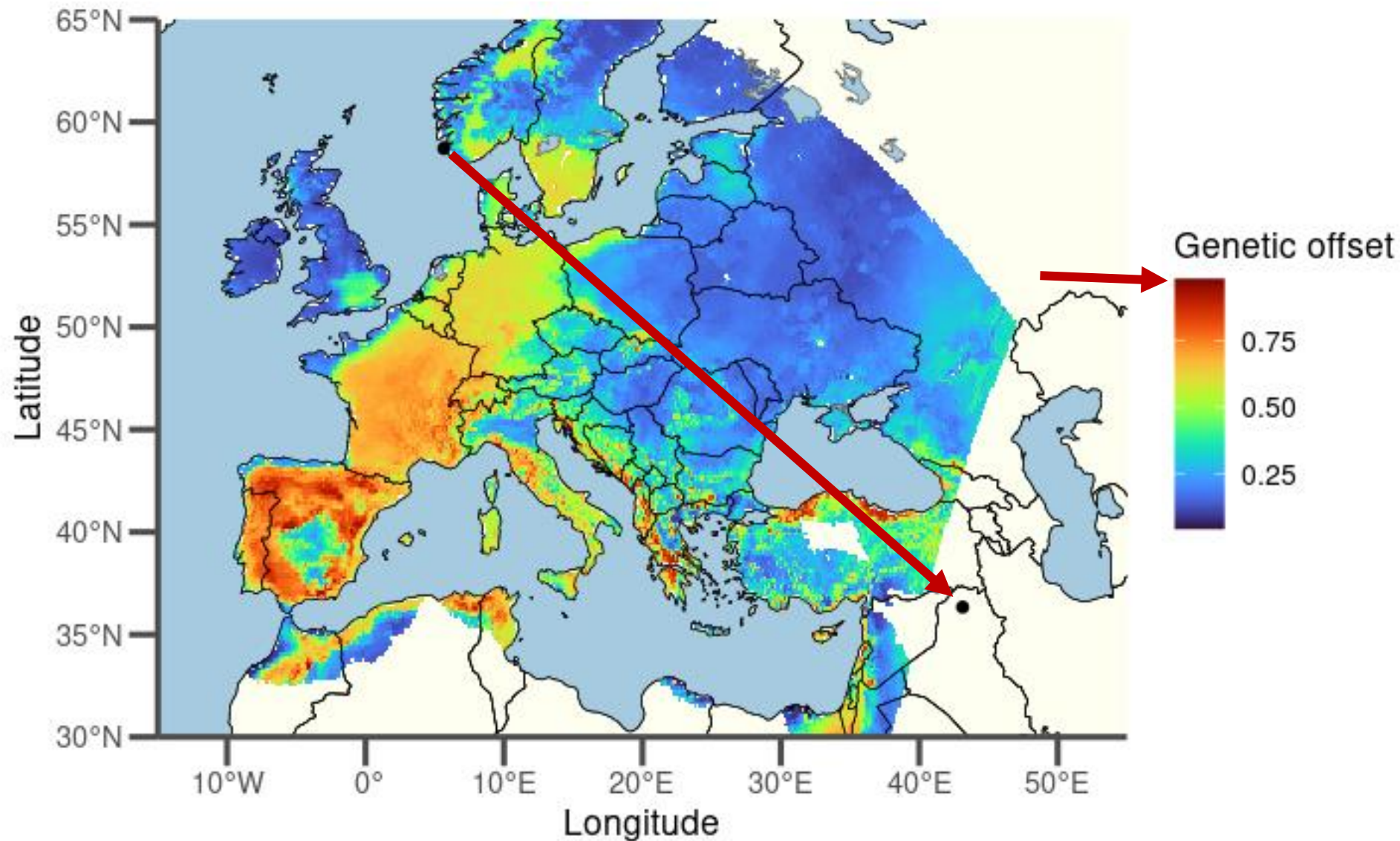
Genetic composition related to adaptation in the current climate

Genetic composition related to adaptation in the future climate RCP 8.5ccl



## ➤ Que faire de ces projections ?

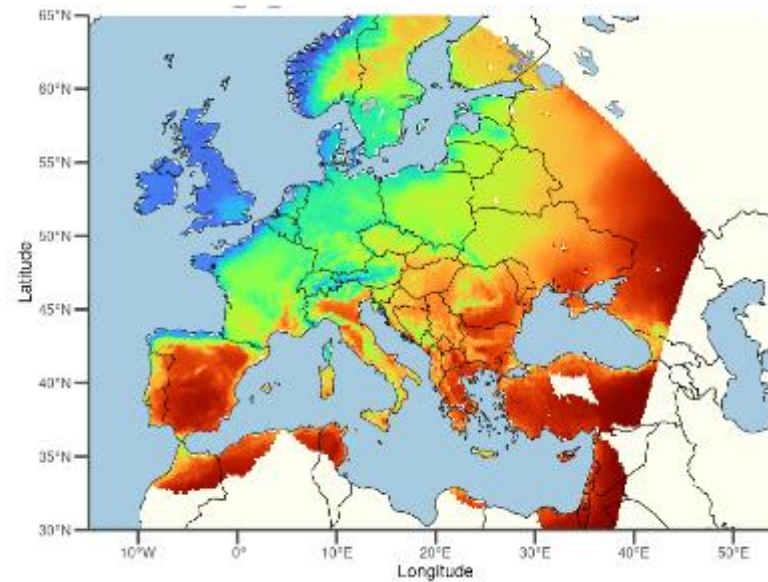
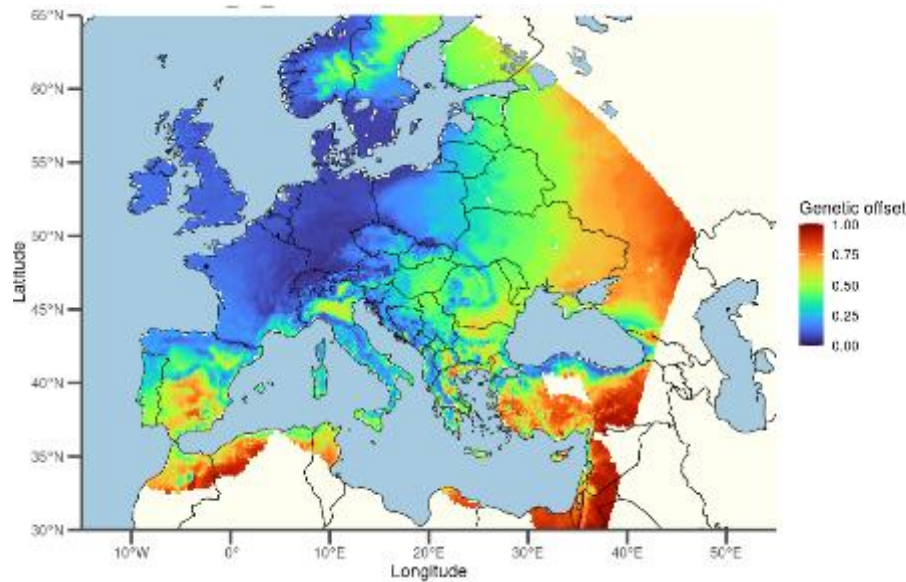
Quelles sont les régions les plus menacées par le climat futur ?



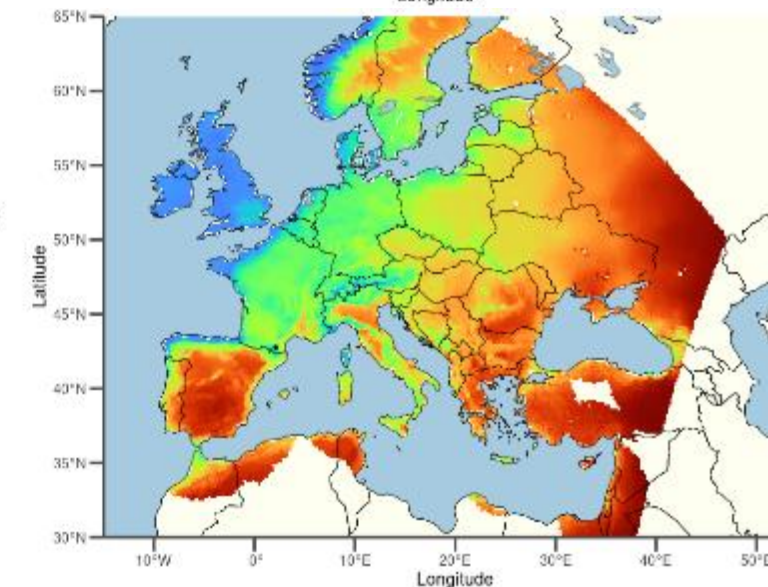
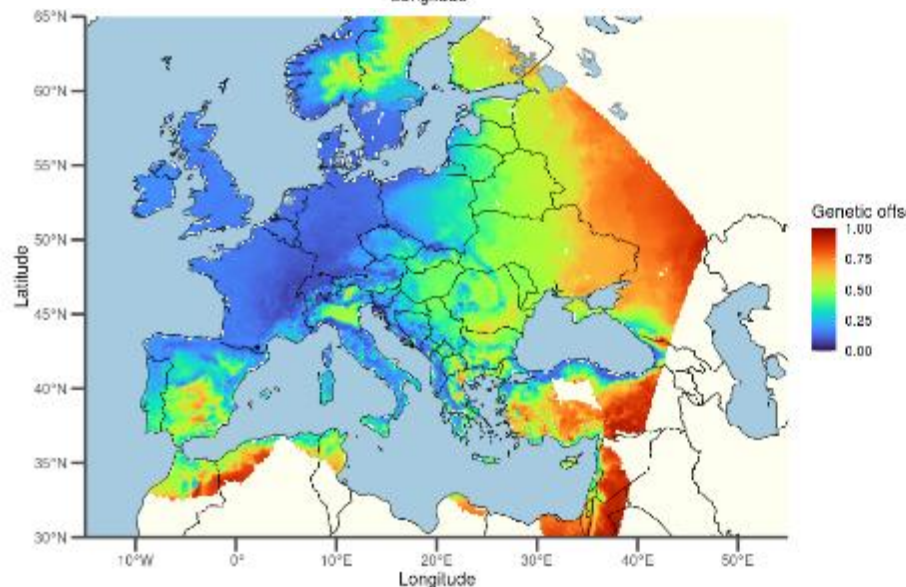


## ➤ Que faire de ces projections ?

Dans quelles régions une population sera-t-elle adaptée au climat futur ?



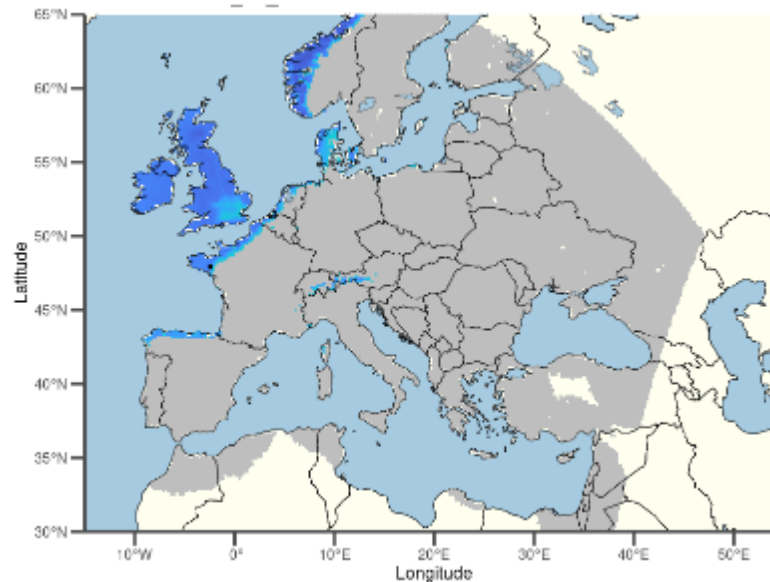
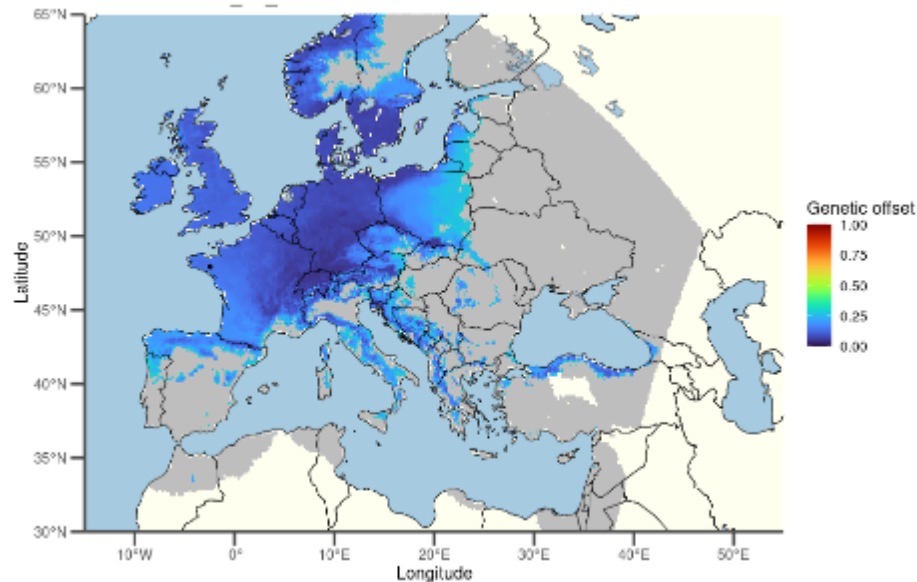
Bretagne (FR 56)



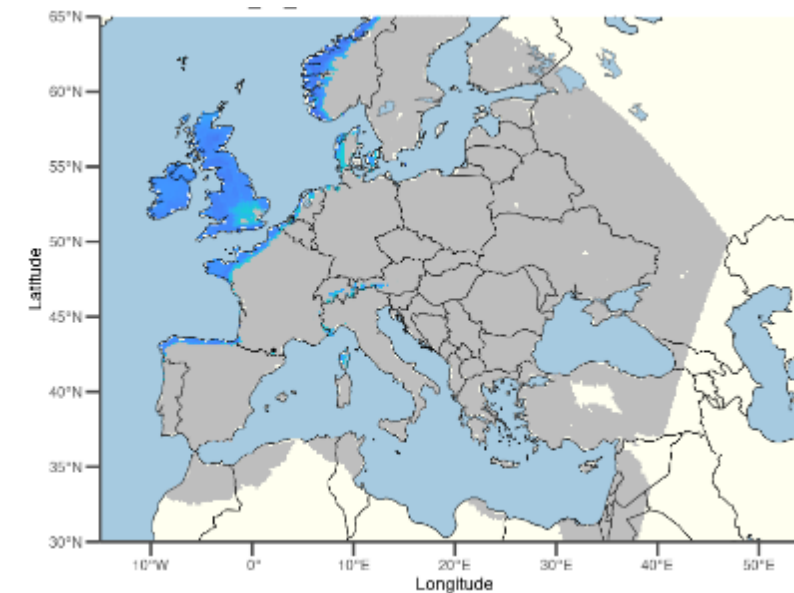
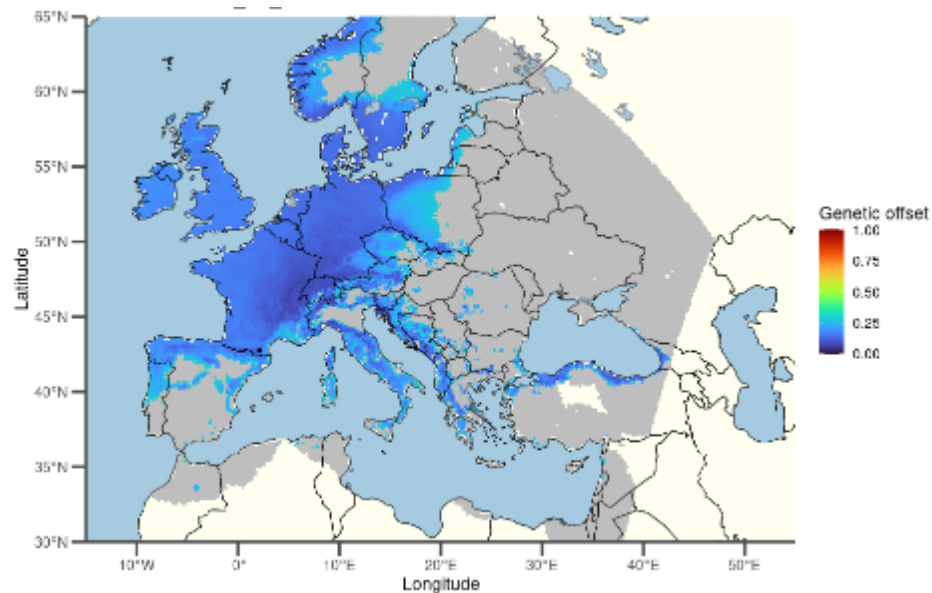
Sud de la France  
(FR 11)

## ➤ Que faire de ces projections ?

Dans quelles régions une population sera-t-elle adaptée au climat futur ?



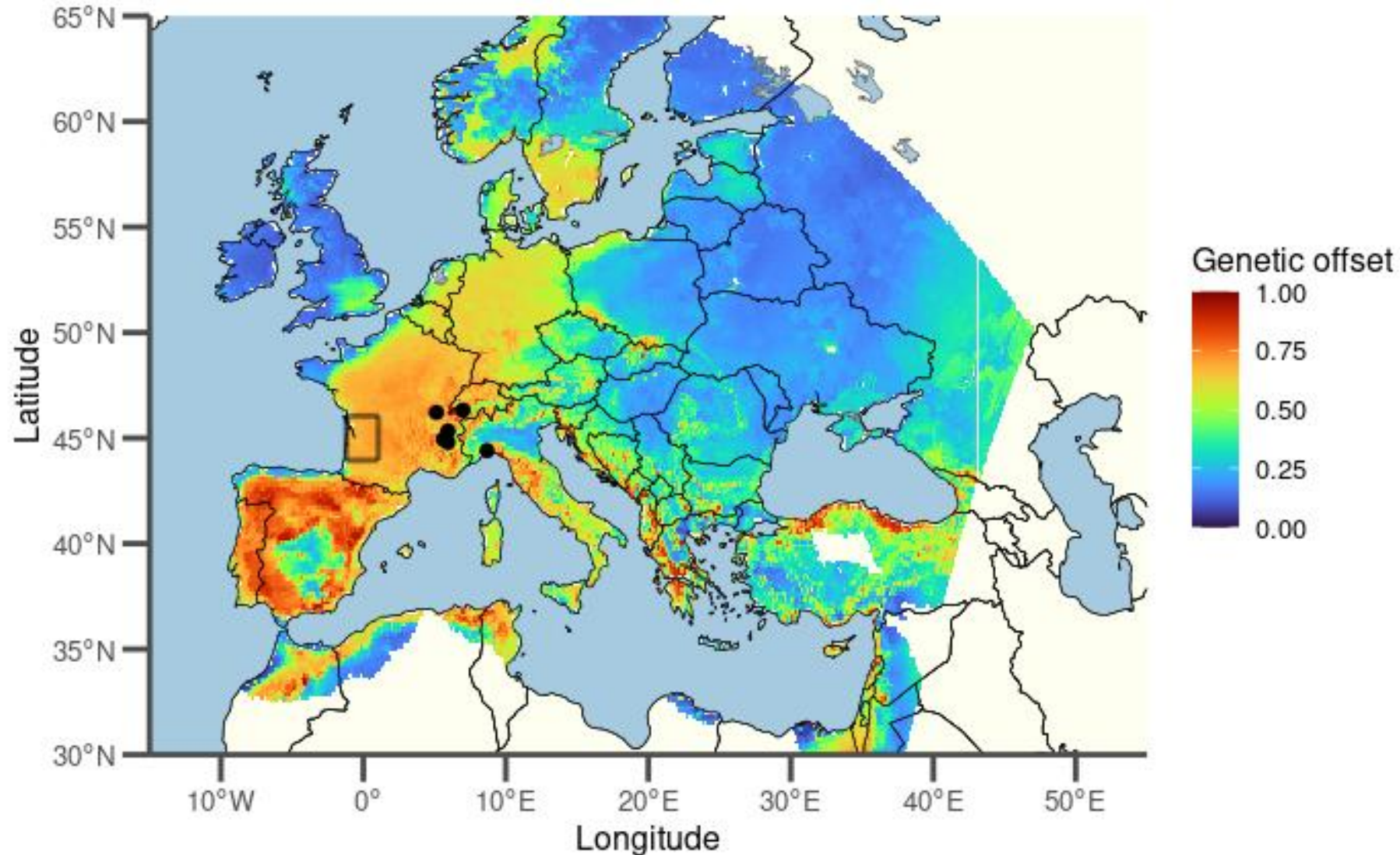
Bretagne (FR 56)



Sud de la France  
(FR 11)

## ➤ Que faire de ces projections ?

Dans ma région, quelle population existante dois-je utiliser pour garantir que les populations de ray-grass de mes prairies sont adaptées à l'avenir ?



# ➤ Que faire de ces projections ?

## Conclusions

En étudiant la vulnérabilité génomique, nous pouvons :

- Identifier les populations en danger Améliorer la gestion des ressources génétiques
- Tenter de déterminer si les populations existantes présentent une préadaptation aux climats futurs prévus.

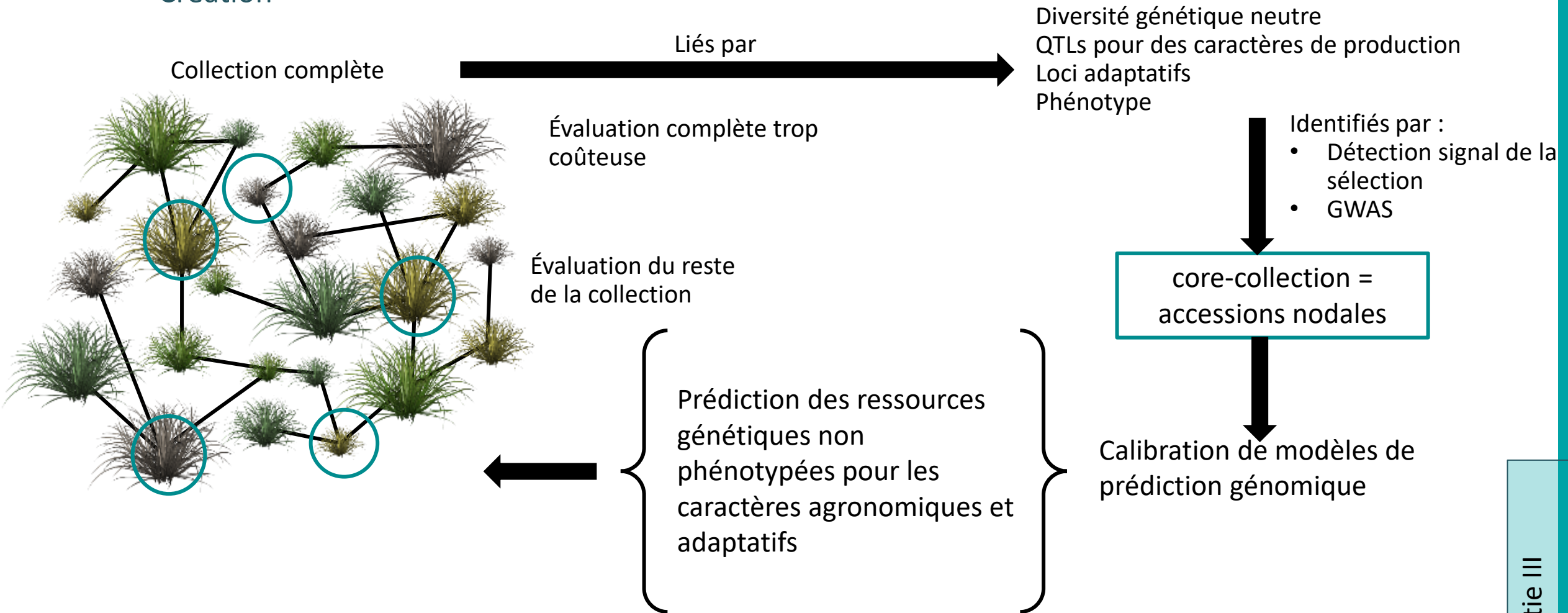
Ces analyses contribueront à :

- Informer sur la migration assistée afin de réduire la vulnérabilité au changement climatique
- Aider les sélectionneurs à identifier les régions où leurs cultivars sont susceptibles de donner les meilleurs résultats.



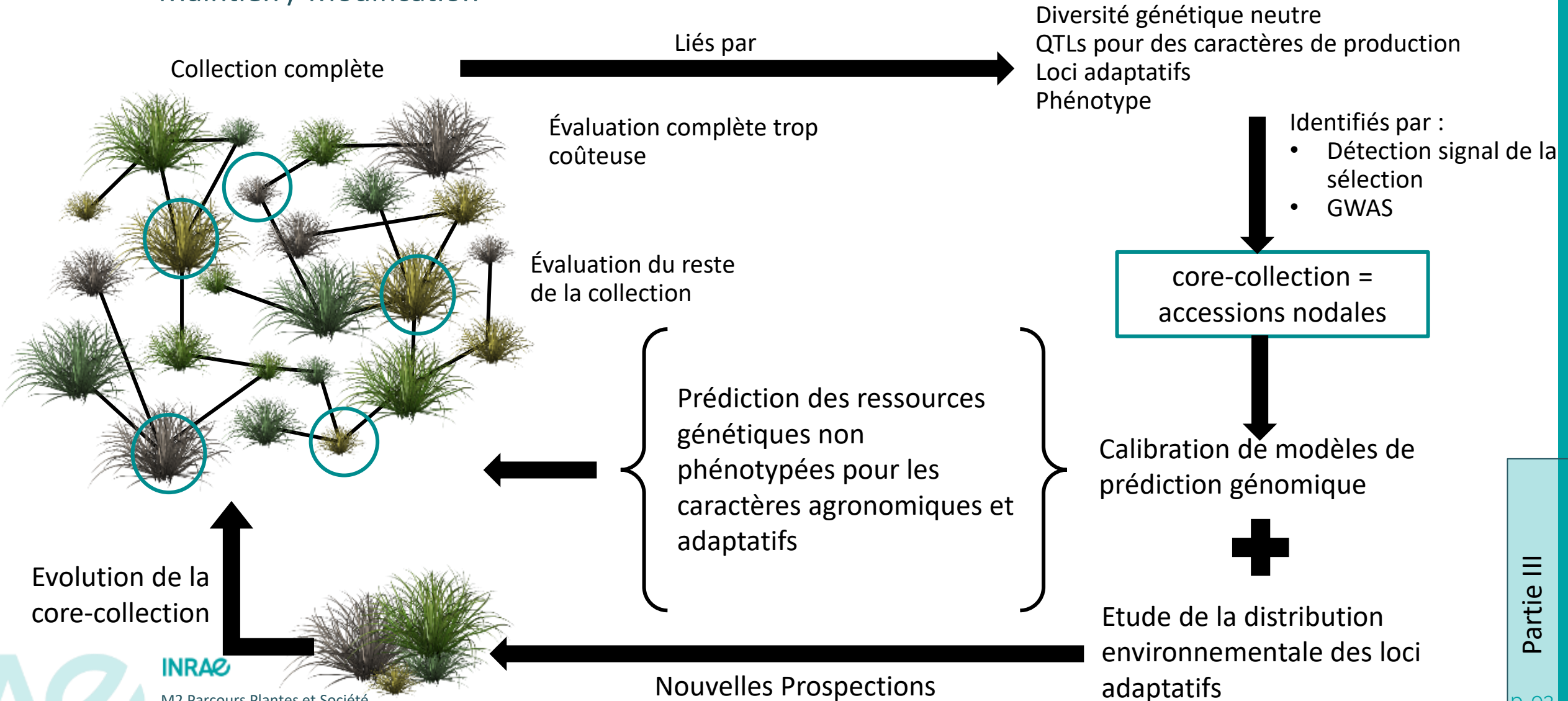
# ➤ Création/maintien d'une core-collection

Création



# ➤ Création/maintien d'une core-collection

Maintien / Modification



INRAE

➤ Conclusions générales



## > Ce qu'il faut retenir

Que ce soit pour l'amélioration des plantes ou la gestion des ressources génétiques

Les marqueurs moléculaires sont un **outils essentiel**

Ils interviennent à **plusieurs moment** et pour **plusieurs utilités**

Ils permettent des **analyses génétiques** plus poussées et plus **détaillées**

Ils permettent l'**accélération** des **programmes d'amélioration** et de la **caractérisation des ressources génétiques**

De part leur coût, Il faut bien réfléchir a :

- Combien
- Quelle méthodologie
- Comment les utiliser
- Pour quelle utilisation





## ➤ Proposition de stage M2

Début mars à fin août 2025

### **Objectifs du Stage :**

1. Vérifier l'impact du pollinarium sur l'émission de pollen.
2. Étudier la relation entre la date d'épiaison et l'émission de pollen.

**Hypothèse :** Les plantes dans le pollinarium, protégées et exposées plein Sud, émettent leur pollen plus tôt que celles situées en dehors de Poitiers.

### **Dispositif Expérimental :**

- Deux sites : Pollinarium de Grand Poitiers et pépinière à Lusignan.
- 15 populations de ray-grass anglais, du très précoce au très tardif.
- Mesures manuelles de mars à août sur la phénologie, morphologie et croissance des plantes.
- Prises de spectre proche infrarouge (NIRS) pour suivre la composition chimique des plantes.

### **Bénéfices :**

- Aider le pollinarium à anticiper l'émission de pollen et prévenir les allergiques.
- Permettre à l'INRAE de mieux comprendre la phénologie du ray-grass anglais.

**INRAE**