

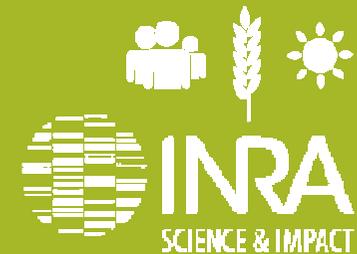


OPTIMISATION DE LA SÉLECTION GRÂCE AUX MARQUEURS MOLÉCULAIRES

Sophie Bouchet
CR2 INRA

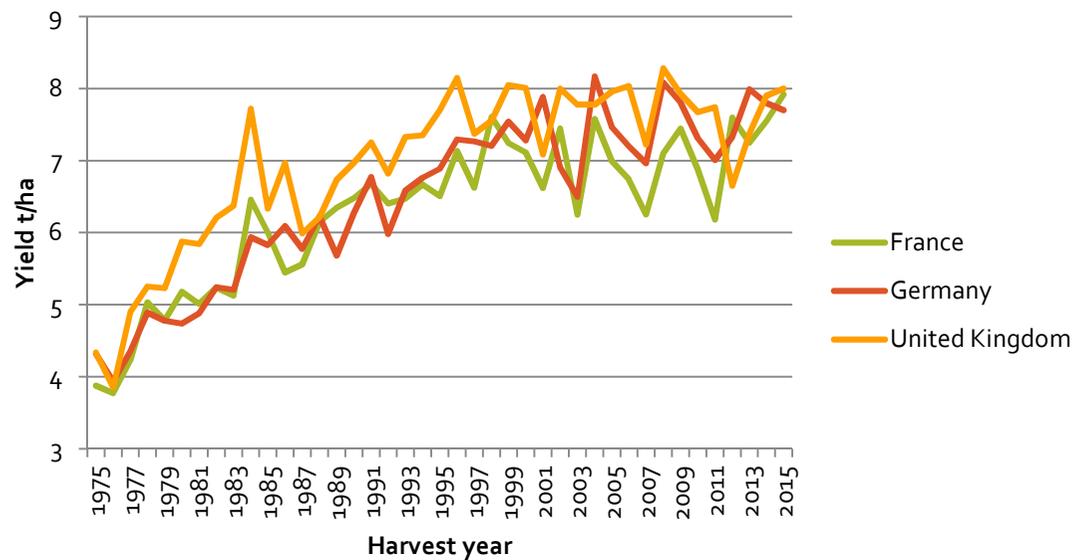
Unité GDEC : Génétique, Diversité et Ecophysiologie des céréales (blé tendre principalement)

Equipe DGS : Diversité Génétique et Sélection



EST-CE QUE LES OUTILS MOLÉCULAIRES PEUVENT AIDER À OPTIMISER LA SÉLECTION VARIÉTALE?

Rendements blé tendre en Europe



LE BLÉ TENDRE

Objectif du sélectionneur:

Pyramider les allèles favorables pour un maximum de caractères dans une même variété

- rendement
- phénologie
- maladies
- stress hydrique et thermique
- utilisation optimale de l'azote
- froid
- verse
- forte teneur en protéines pour l'export
- qualité boulangère pour le marché intérieur



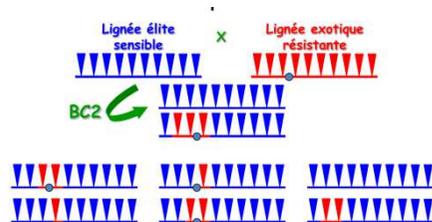
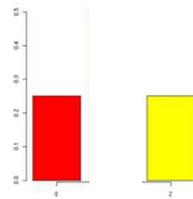
classer les lignées le plus tôt possible dans le programme de sélection grâce

- observations / notations au champ
- marqueurs diagnostiques (on connaît l'allèle favorable pour un gène majeur qui contrôle le caractère qui nous intéresse ou des scores moléculaires)

On veut optimiser gain génétique / gain économique

Objectif : classer les lignées

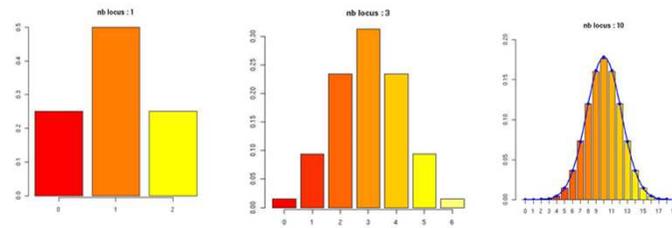
Caractères simples contrôlés par gènes majeurs



BC, OGM, CRISPR

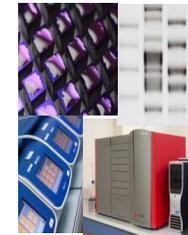
Marqueurs diagnostiques

Caractères complexes



Sélection génomique
Génétique quantitative

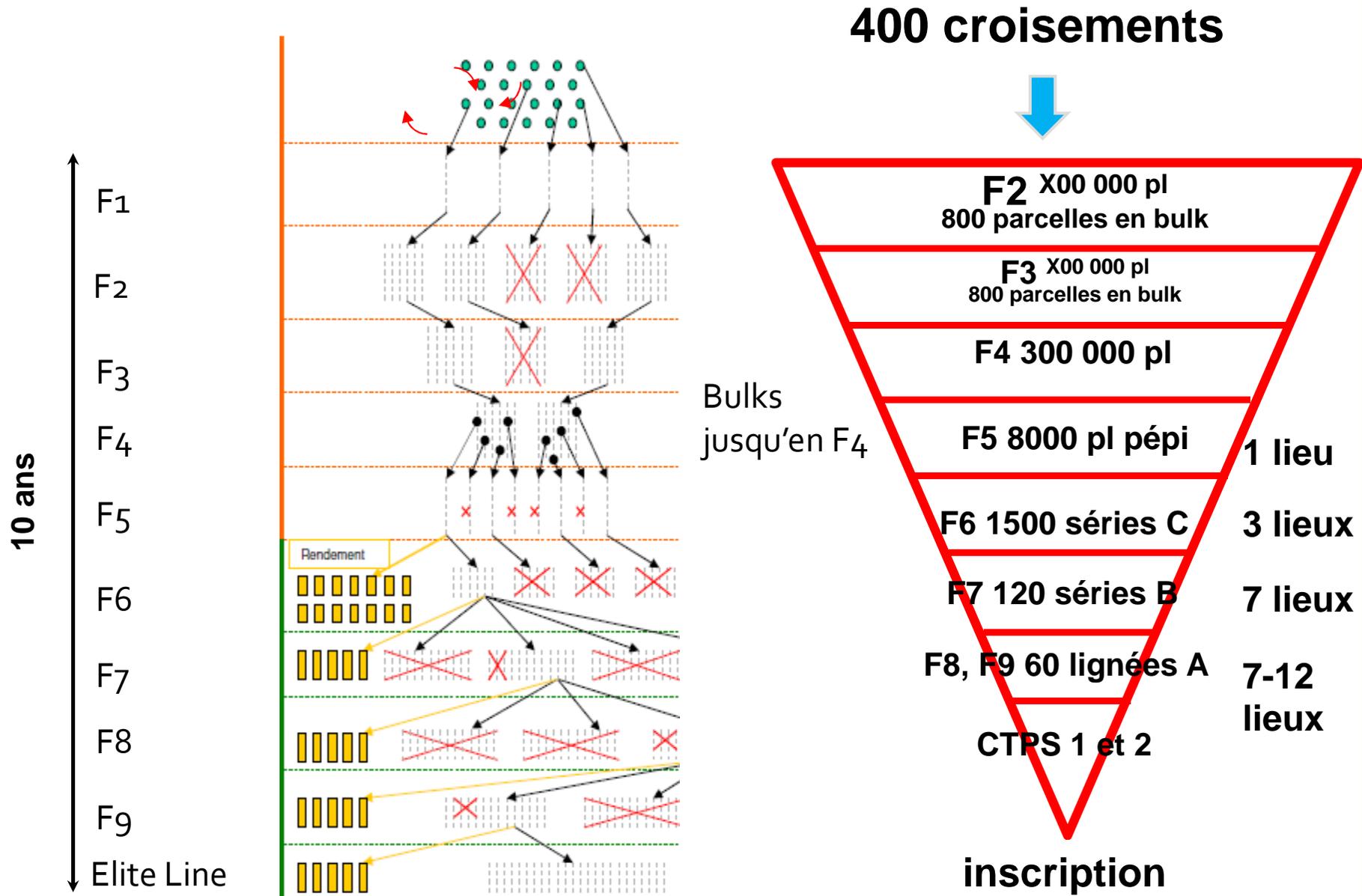
Équation de prédiction



Score moléculaire: pour chaque lignée,
pour chaque croisement

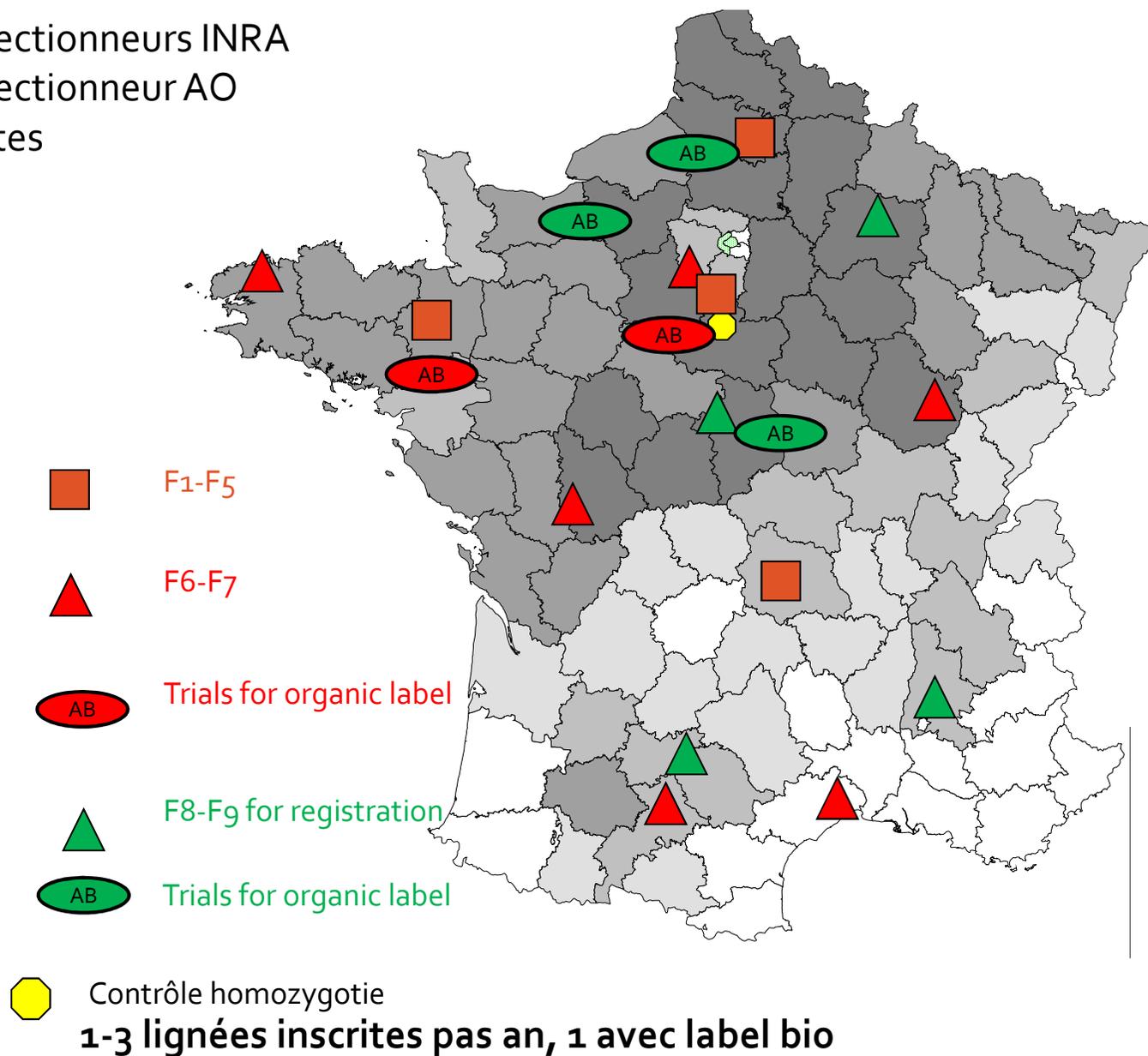
LE PROGRAMME DE SÉLECTION INRA-AGRIOBTENTIONS

Schéma de sélection classique blé



Réseau d'essais INRA-AgriObtentions

3 sélectionneurs INRA
1 sélectionneur AO
15 sites



Inscription

| Classes qualité | Note de panification | Rendement |
|--------------------------------|----------------------|----------------|
| Blé Panifiable Supérieur (BPS) | note > 250 | > 102% témoins |
| Blé Panifiable (BP) | 220 < note < 250 | > 104% témoins |
| Blé Autre Utilisation (BAU) | note < 220 | > 107% témoins |

Base de données de 20 ans de phénotypage du programme de sélection (F8-F9)

| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 2000 | 15 | | | | | | | | | | | | | |
| 2001 | 13 | 28 | | | | | | | | | | | | |
| 2002 | 7 | 13 | 22 | | | | | | | | | | | |
| 2003 | 2 | 3 | 10 | 37 | | | | | | | | | | |
| 2004 | 1 | 2 | 4 | 8 | 32 | | | | | | | | | |
| 2005 | 1 | 1 | 2 | 3 | 9 | 24 | | | | | | | | |
| 2006 | 1 | 1 | 1 | 1 | 6 | 12 | 38 | | | | | | | |
| 2007 | 1 | 1 | 2 | 2 | 5 | 6 | 17 | 58 | | | | | | |
| 2008 | 1 | 1 | 1 | 1 | 3 | 2 | 6 | 21 | 47 | | | | | |
| 2009 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 5 | 18 | 44 | | | | |
| 2010 | 1 | 1 | 1 | 1 | 3 | 2 | 3 | 4 | 8 | 16 | 55 | | | |
| 2011 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 4 | 5 | 17 | 48 | | |
| 2012 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 3 | 6 | 19 | 55 | |
| 2013 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 3 | 3 | 11 | 29 | 75 |

RESSOURCES GÉNÉTIQUES

Collection de blé tendre



12 000 accessions

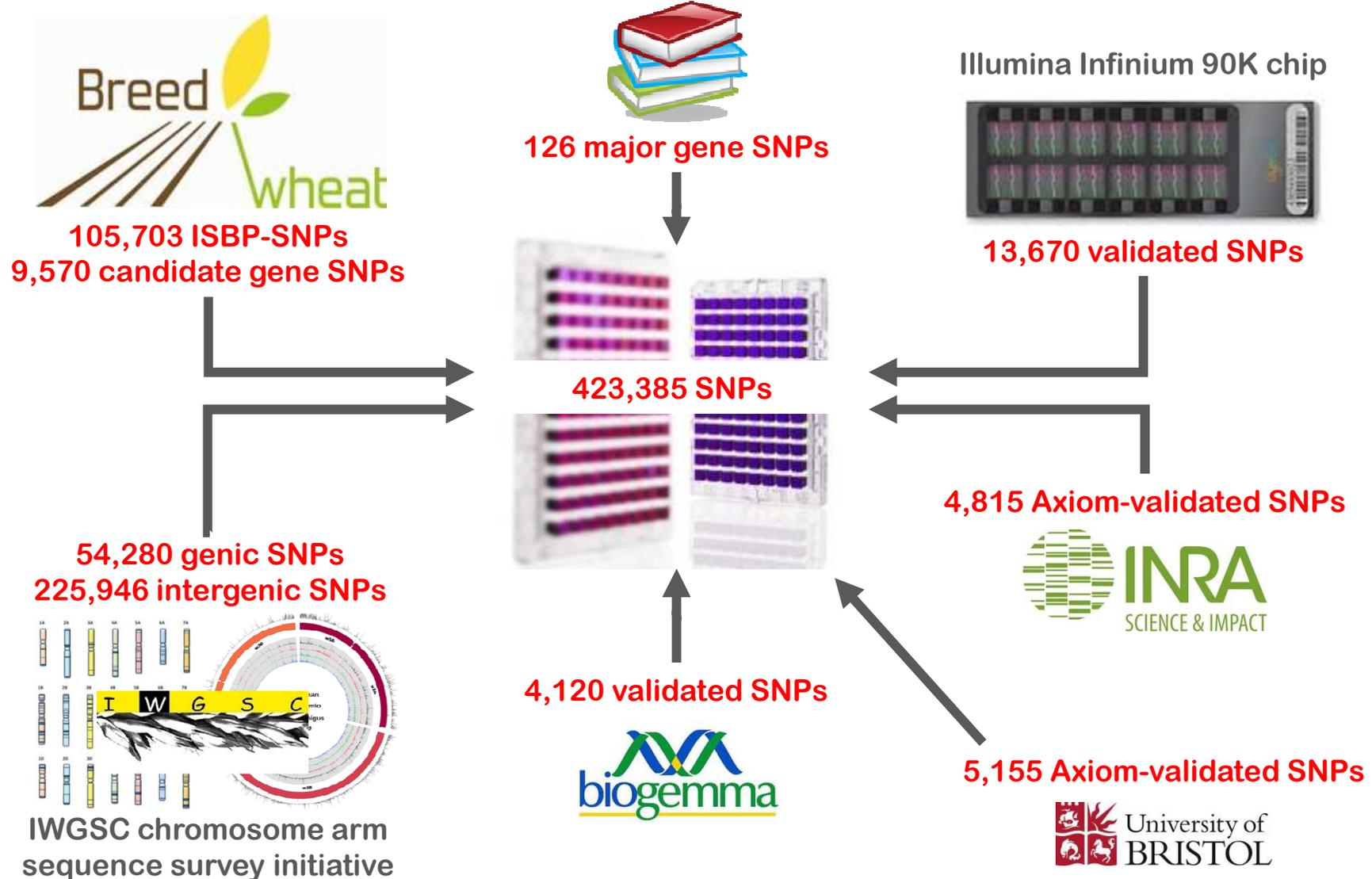


Triticum

- Landraces
- Variétés traditionnelles (inscrites avant 1960)
- Lignées modernes (inscrites depuis 1960)
- Phénotypage Breedwheat

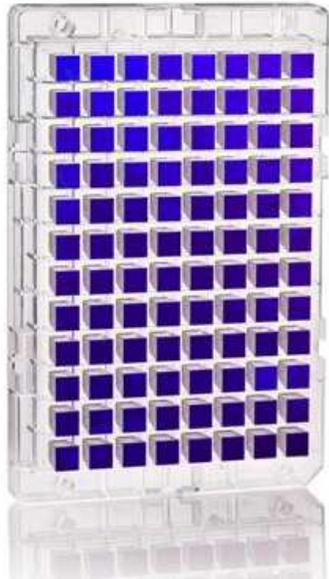
PUCES DE GÉNOTYPAGE

puce TaBW420K / puce TaBW180K



Rimbert et al, Paux, 2018

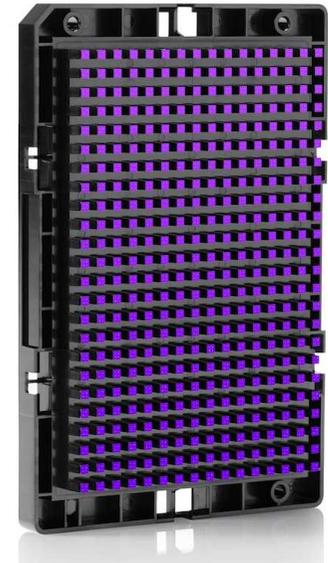
Construction de la puce TaBW35K



TaBW420K array
423K SNPs x 96 samples

130 euros /
individu

- ✓ Publics
- ✓ Répartition uniforme le long de la carte génétique
- ✓ PolyHighRes ou OTV
- ✓ Position physique sur IWGSC RefSeq v1.0
- ✓ En commun avec Bristol35K
- ✓ Associés avec 1 caractère
- ✓ Bon critère de qualité de Génotypage (profil de fluorescence qui séparent bien les 2 allèles)

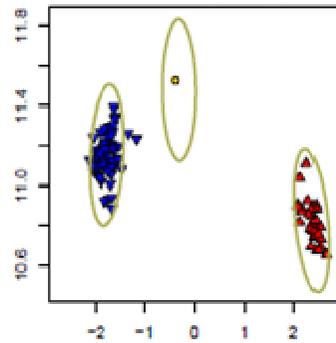


TaBW35K SNP array
35K SNPs x 384 samples

35 euros / individu

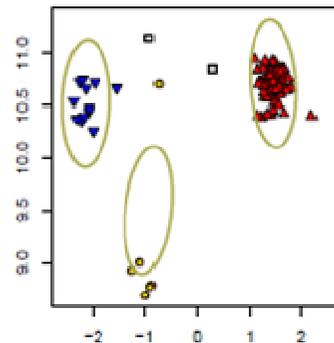
Classes de qualité des marqueurs

PolyHighResolution



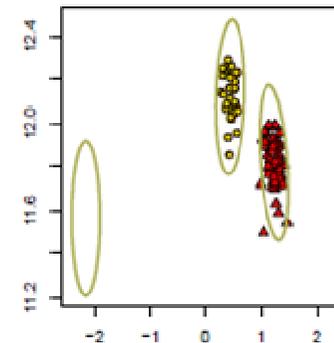
246,962
58.3%

OTV



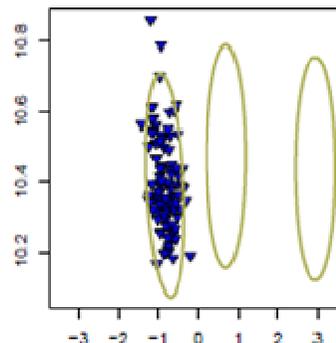
37,712
8.9%

No Minor Hom



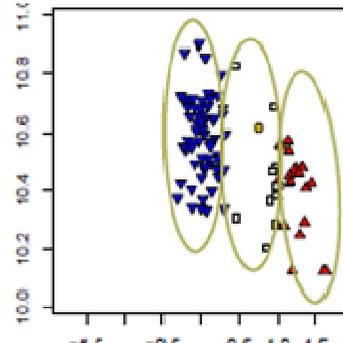
15,305
3.6%

MonoHighResolution



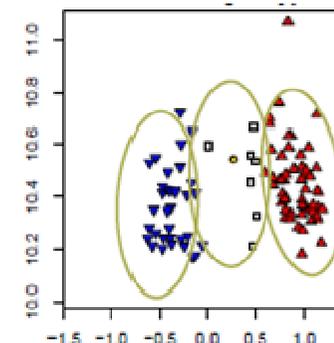
80,988
19.1%

CallRate Below Threshold



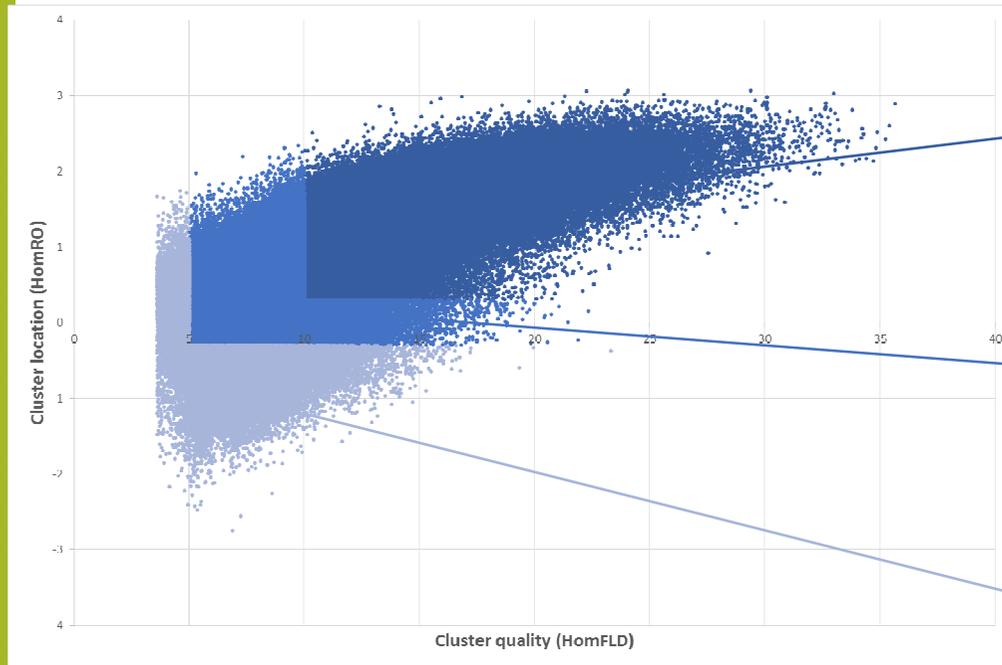
22,633
5.3%

Other

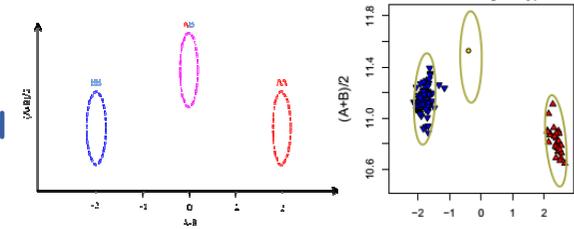


19,785
4.7%

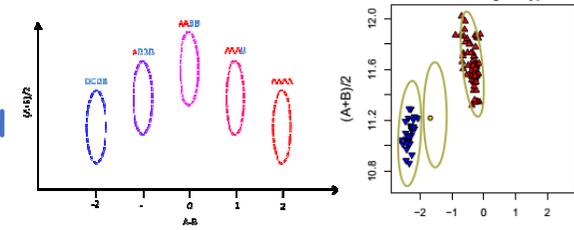
Classes de qualité des marqueurs



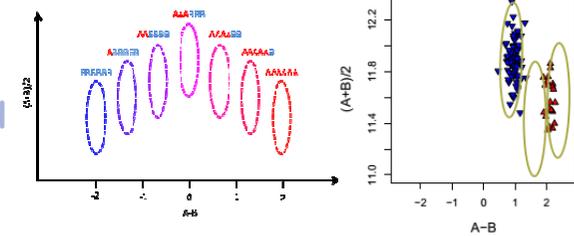
Diploid



Tetraploid



Hexaploid

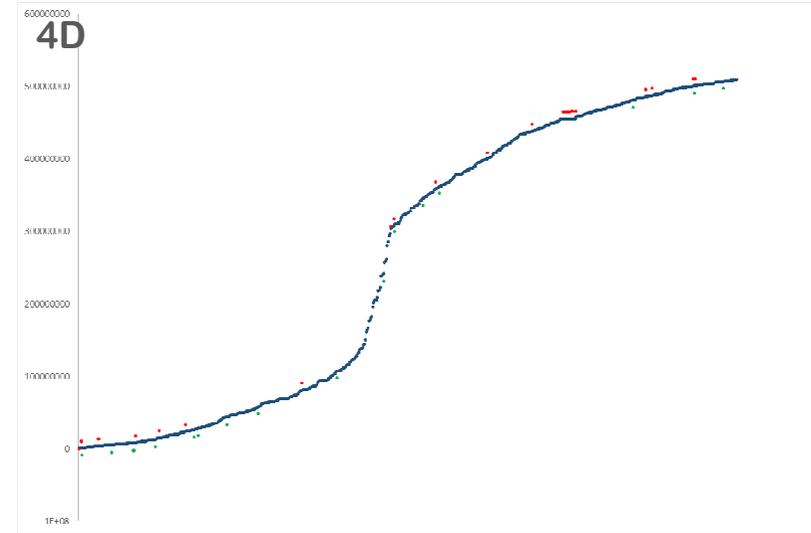
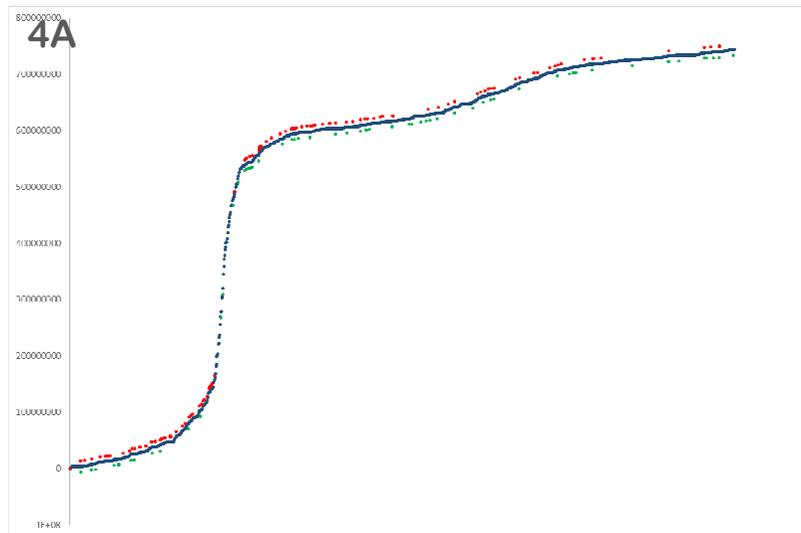
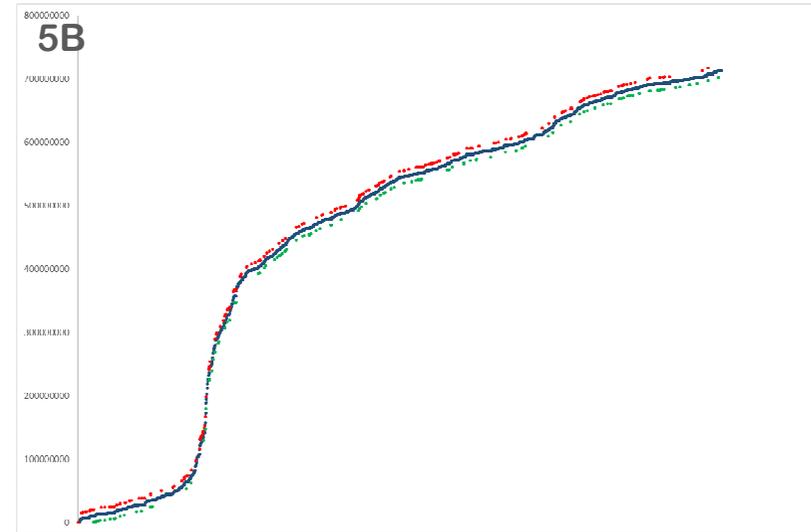
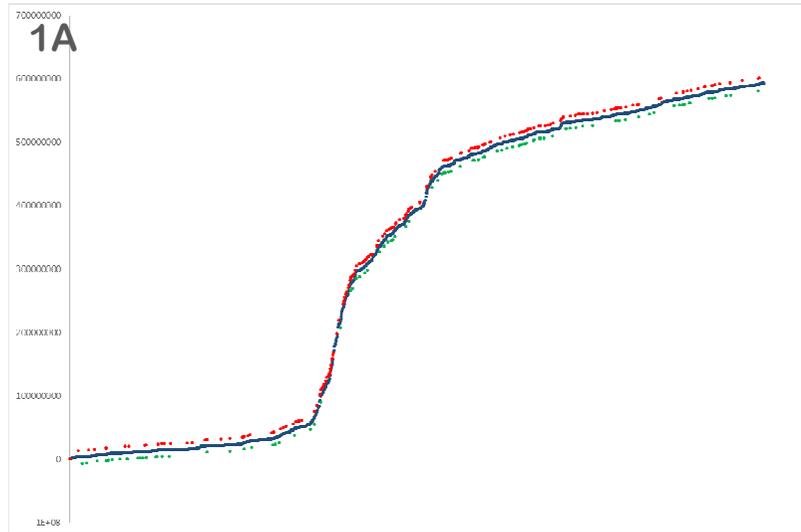


→ 141,246 (49.6%) diploidized

111,892 (39.3%) tetraploidized

31,537 (11.1%) hexaploidized

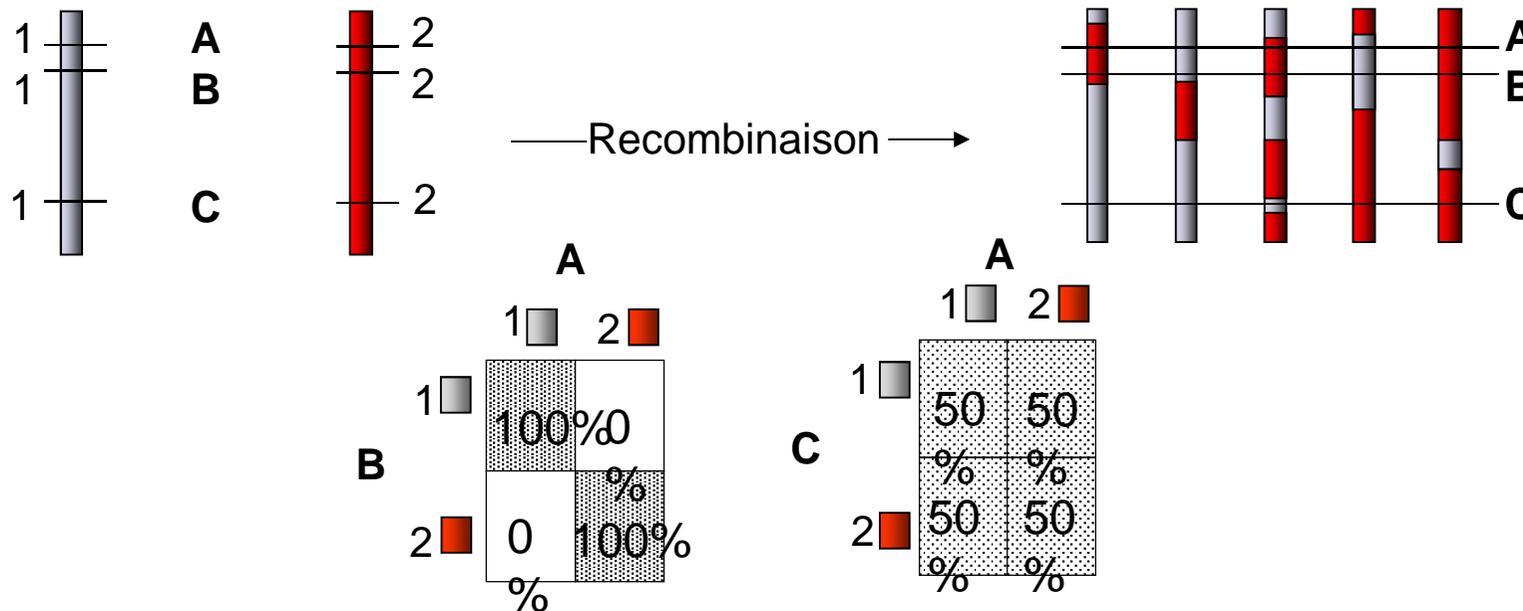
Distribution des SNPs qui ne sont pas en Déséquilibre de liaison ($r^2 < 0.8$, LD pruning)



TaBW35K SNPs shared with **Bristol35K** or **Illumina 90K**

Déséquilibre de liaison : définition

- A, B, C sont 3 loci avec 2 allèles chacun notés 1 and 2



- DL = association non aléatoire des allèles à différents loci
- DL = écart à l'indépendance statistique des fréquences alléliques à différents loci
- $DL = P(A1B1) - P(A1) \cdot P(B1) = P(A2B2) - P(A2) \cdot P(B2)$

420K vs 35K : enrichissement en OTV et marqueurs diploïdes

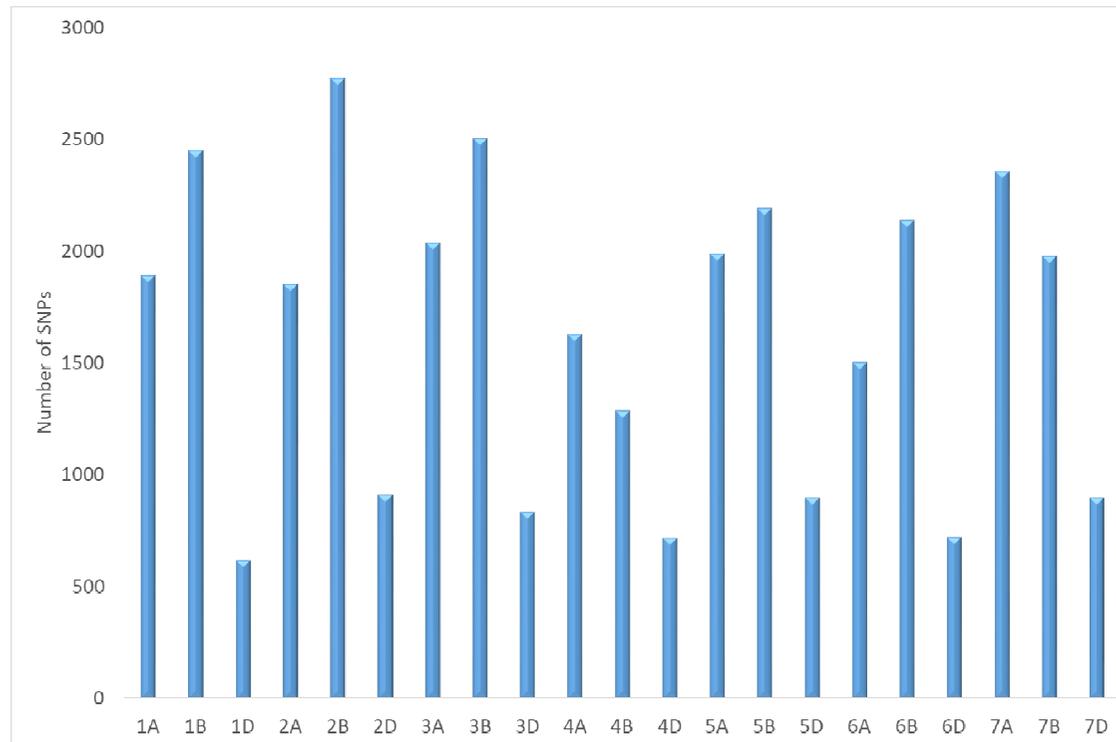
| | PHR | OTV |
|------------|-----|-----|
| 232645 SNP | 86% | 14% |
| 34746 SNP | 75% | 25% |

| | diploid | tetraploid | hexaploid |
|------------|---------|------------|-----------|
| 232645 SNP | 53% | 37% | 10% |
| 34746 SNP | 59% | 34% | 6.5% |

puce TaBW35K

- ✓ 34 746 markers
- ✓ 25 916 PHR
- ✓ 8830 OTV
- ✓ 3229 commun 35K bristol
- ✓ 2190 associés à un caractère (biogemma)

Puce TaBW35K



✓ A-genome

- ❖ 14,180 SNPs
- ❖ 2,116 SNPs / chromosome
- ❖ One SNP every 378 kb

✓ B-genome

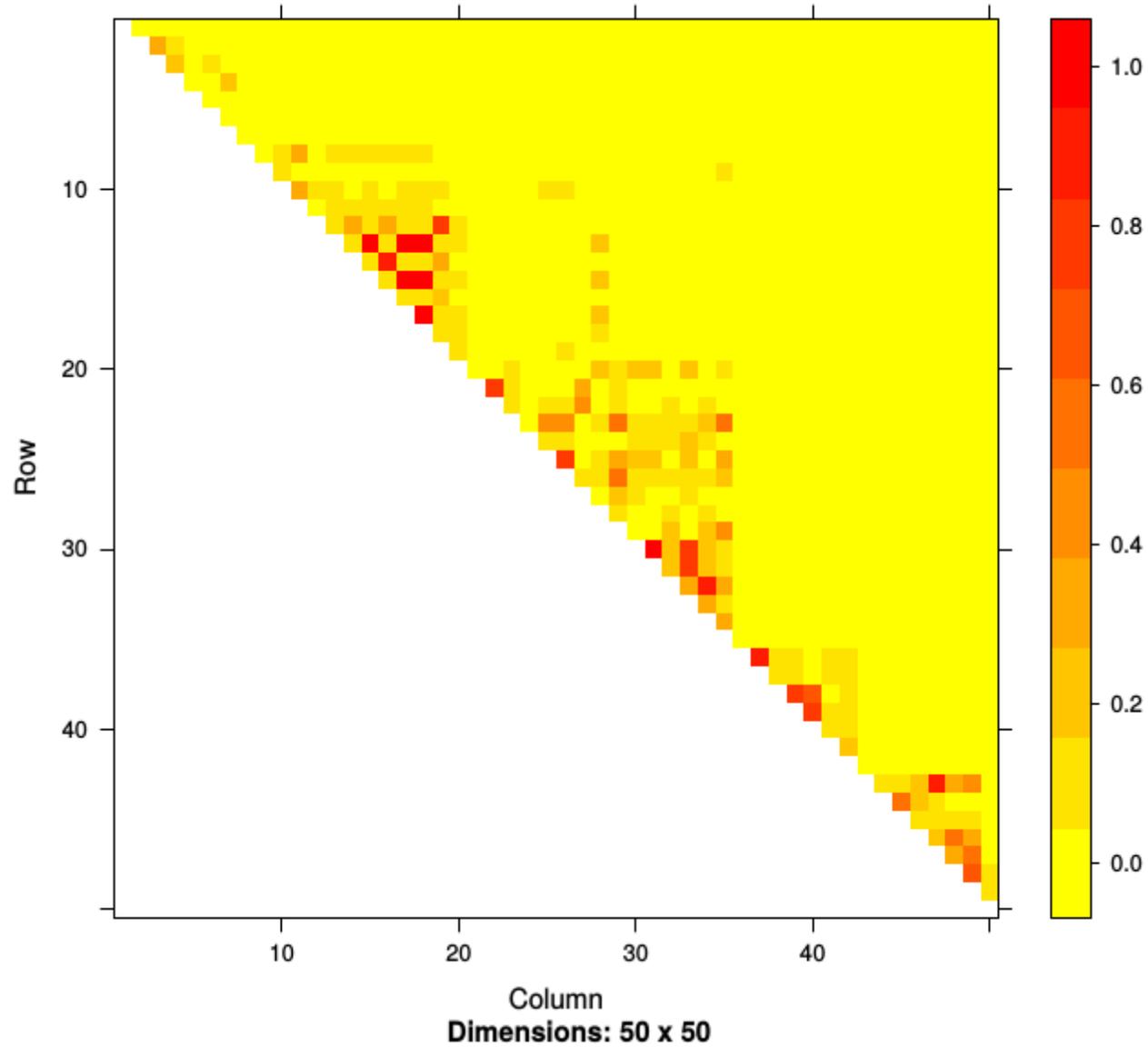
- ❖ 15,414 SNPs
- ❖ 2,202 SNPs per chromosome
- ❖ One SNP every 356 kb

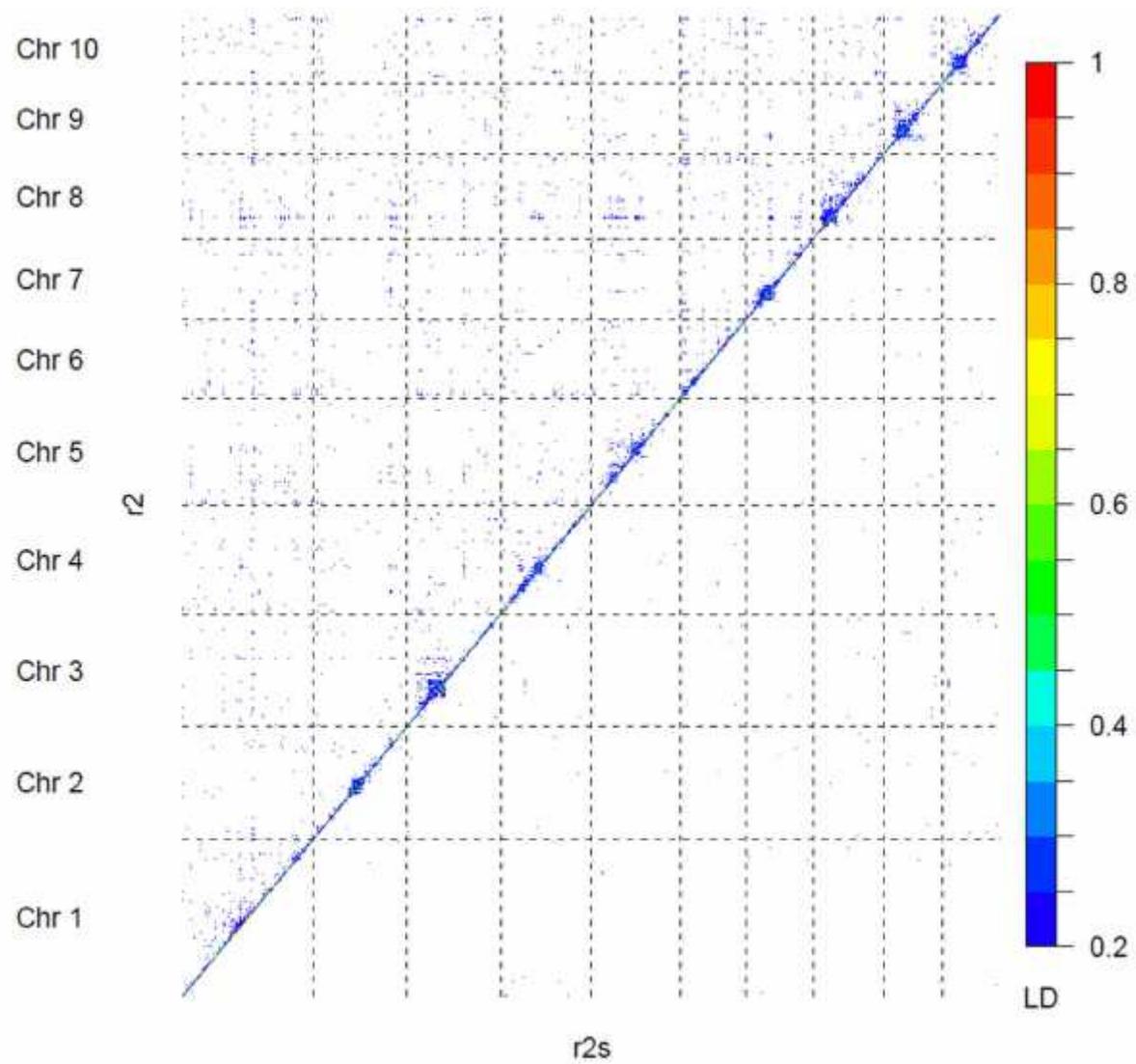
✓ D-genome

- ❖ 9,383 SNPs
- ❖ 1,340 SNPs per chromosome
- ❖ One SNP every 525 kb

CONSTRUCTION D'HAPLOTYPES

Visualisation des blocs haplotypiques le long du génome





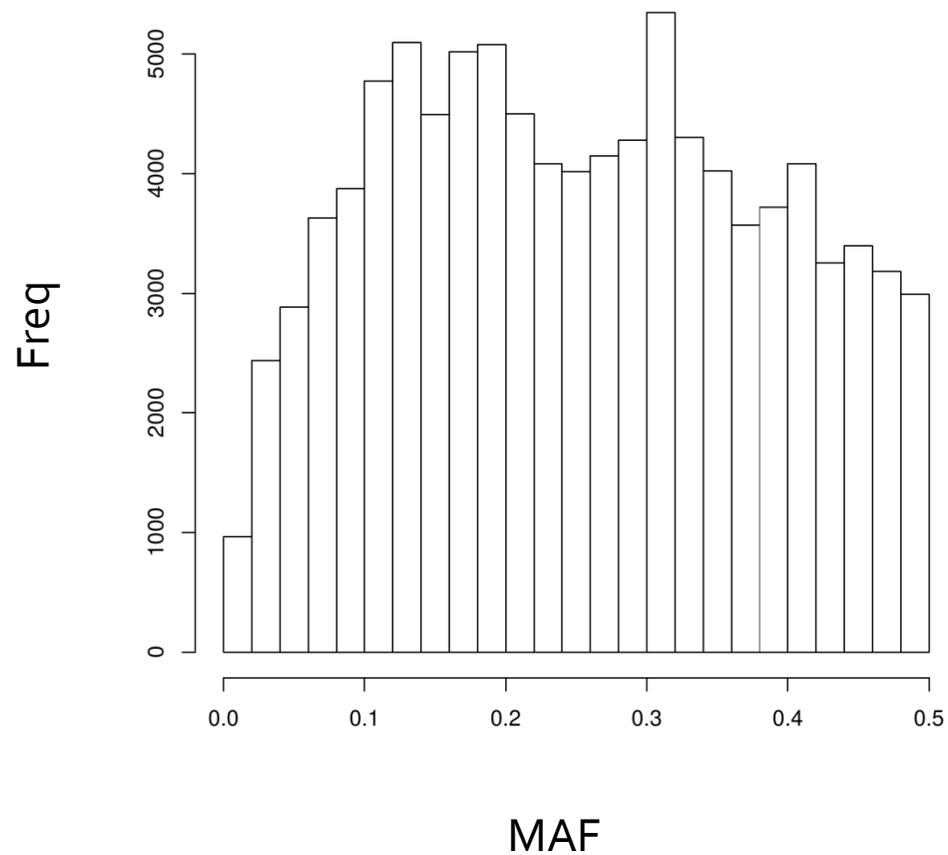
Maïs, bouchet et al, (2013)

Définition des blocs haplotypiques

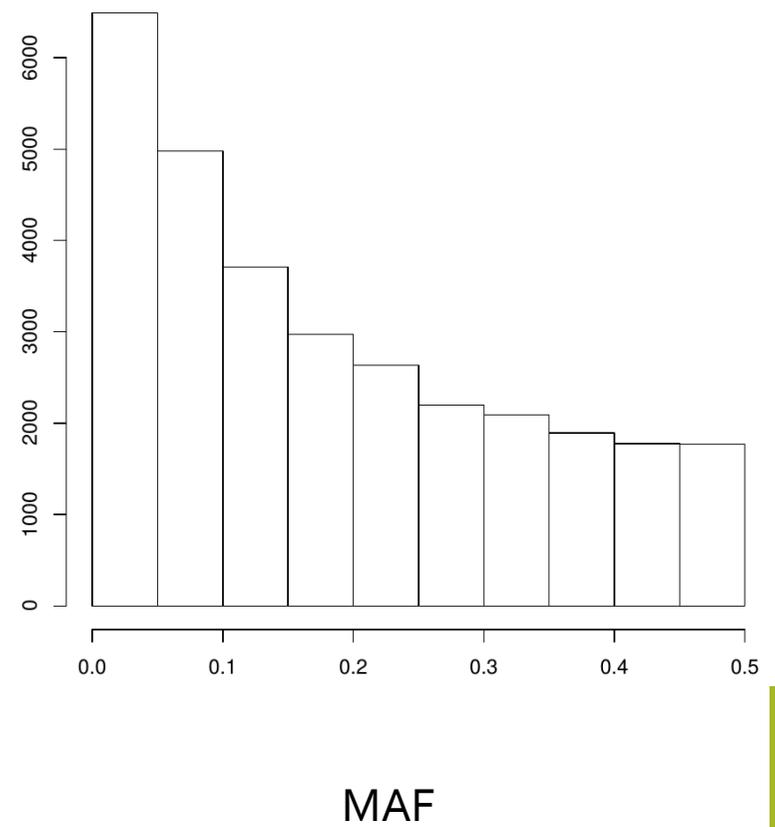
- 95% des valeurs de DL entre 2 marqueurs (D') sont supérieures à 0.98.
- Sur 4403 lignées représentatives de la diversité mondiale (panel WP3),
8741 blocs haplotypiques
- Nb haplotypes par chr corrélé à la taille du chr
80% < 1Mb
- Taille des blocs corrélée au profil de DL et de recombinaison
- Entre 2 et 20 allèles (4 en moyenne) (cf SSR)

Distribution MAF SNP / haplotypes

SNP

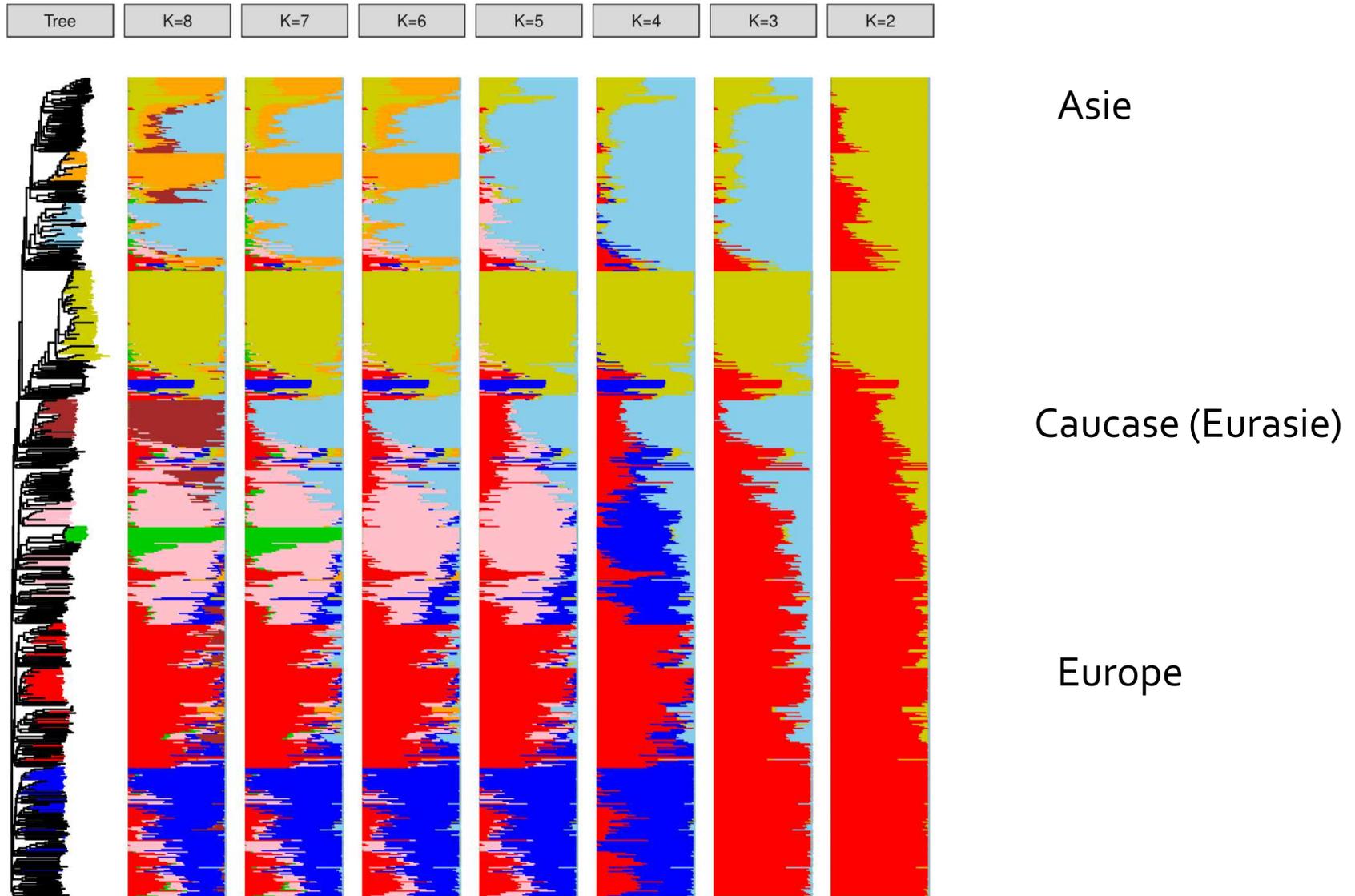


haplotypes



CONSTRUCTION D'EMPREINTES GÉNÉTIQUES

Empreinte génétique de 7000 lignées (élites, landraces, MAGIC, NAM, AB-QTL, programme de sélection)



Perspective Chromosome painting

si on considère qu'il y a 8 groupes ancestraux, quel fragment du génome provient de quel ancêtre?

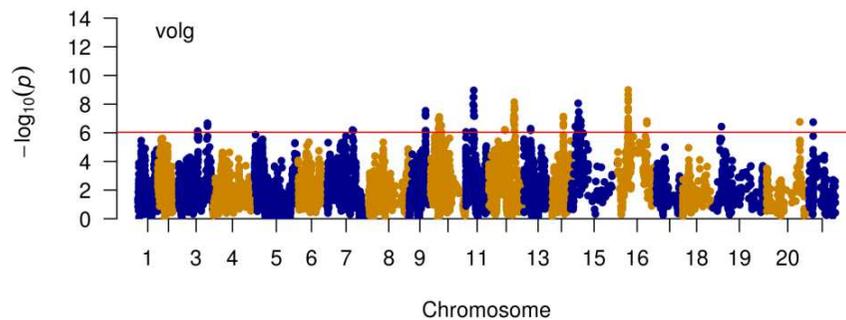
lignées



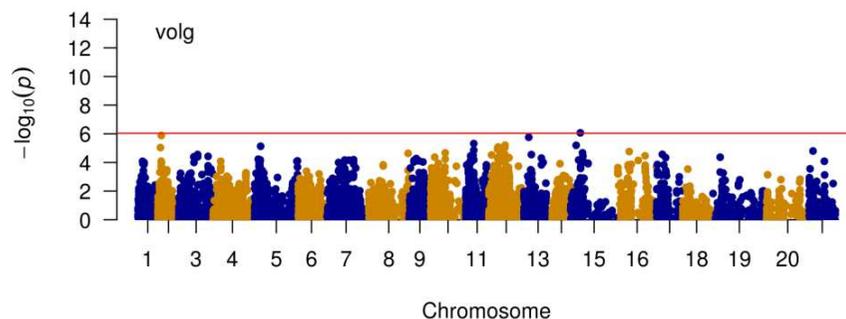
Position du marqueur (Jakobsson et al, 2008)

Perspective : augmenter la puissance de détection des QTLs (études d'association)

Manhattan Plot-core SNP



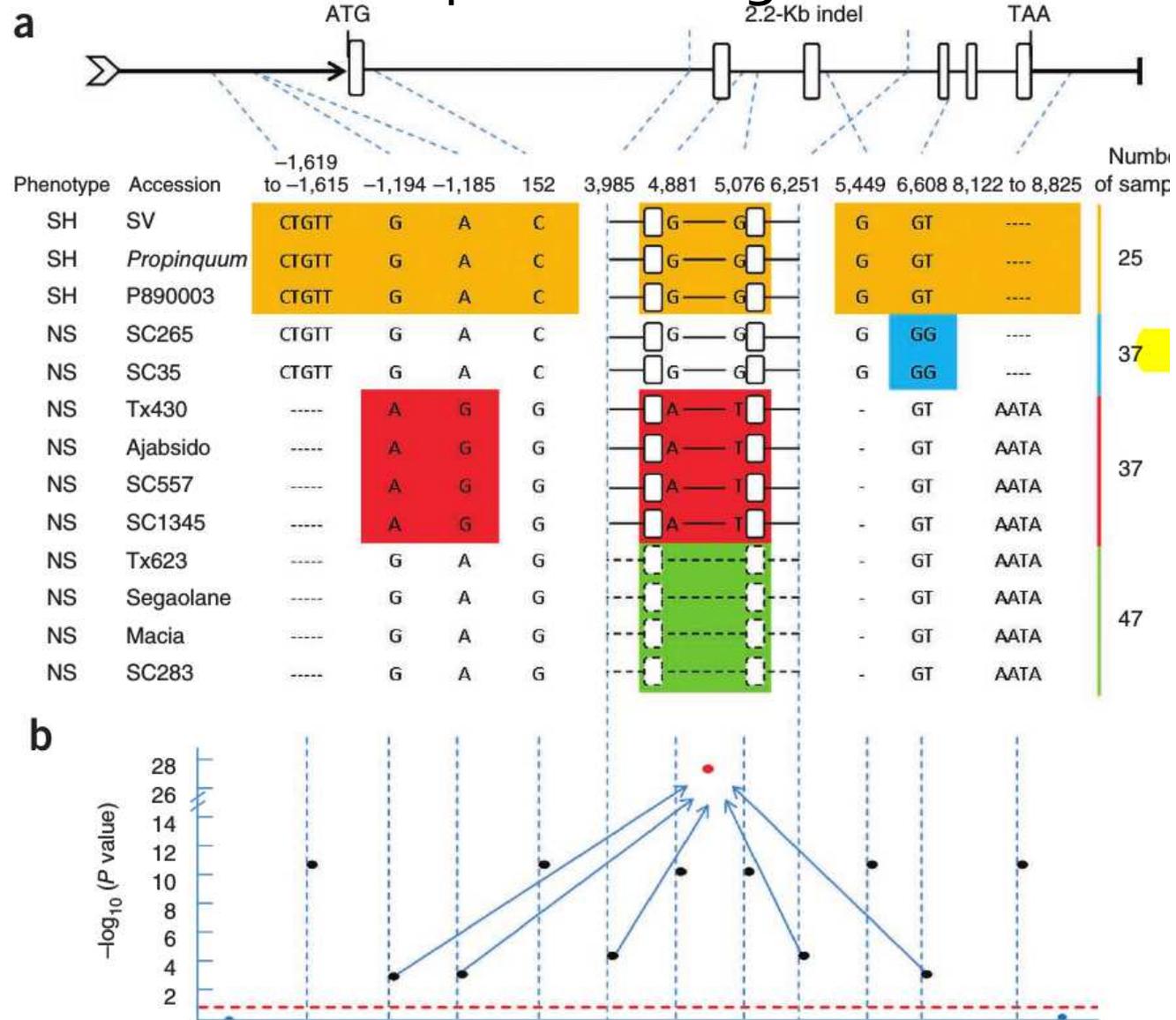
Manhattan Plot-single SNP



| Trait | $-\log_{10}$ (Psingle) Best hit | $-\log_{10}$ (Phap) Best hit | Nb QTLs single | Nb QTLs hap |
|-------|---------------------------------------|------------------------------------|-------------------|----------------|
| npani | 4.9 | 6.7 | 0 | 1 |
| npat | 8.2 | 7.8 | 2 | 19 |
| volg | 6.1 | 9 | 1 | 41 |

1. Modélisation des haplotypes locaux (modèle HMM) et apparentement local
2. Régression haplotypique (modèle bayésien)

Exemple Sh1 Sorgho



Objectif: le faire de manière automatique sur l'ensemble du génome, sans définition de fenêtre

Lin et al, 2012

DÉVELOPPEMENT ET GÉNOTYPAGE EN ROUTINE DE MARQUEURS DIAGNOSTIQUES

Caractères mono géniques

Puce fluidigm 96 marqueurs
10 euros / individu

81 Kaspar (dans les gènes ou en DL)

| Trait | Nb marker | Locus | Specialist | Project | Controls + | Controls - |
|----------------------------|------------------------|-------|-------------------|-------------|--|--|
| vrn | 11 | | jonathan kitt | BW | | |
| rht | 2 | | jonathan kitt | BW | | |
| ppd | 9 | | jonathan kitt | BW | | |
| mosaic | 1sbm1 | | | FSOV SAM-BD | | |
| fusa | 3fhb1 | | | FSOV SAM-BD | | |
| leaf rust | 9Lr 21, 34, 37, 46, 51 | | | FSOV SAM-BD | | |
| yellow rust | 1Yr36 | | | FSOV SAM-BD | | |
| stem rust | 2Sr 2, 36 | | | FSOV SAM-BD | | |
| eyespot (piétain verse) | 1pch1 | | | FSOV SAM-BD | BERMUDE, RYTMIC (tolérant) | APACHE, ALTIGO (sensible) |
| drought tolerance | 21RS1AL, 1RS1BL | | | FSOV SAM-BD | | |
| glutenins | 15 | | catherine ravel | | | |
| PHS (germination sur pied) | 6vp1 | | | FSOV SAM-BD | | |
| Hardness (dureté) | 2pinA-D1, pinB-D1 | | catherine ravel | | | |
| protein | 1gpc1 | | | FSOV SAM-BD | | |
| leaf rust | 4Lr16 | | cyrille saintenac | | BW278, AC Majestic, AC Domain, Kenyon, McKenzie | AC Foremost, Glenlea, RL4452, 86ISMN 2137, Thatcher |
| leaf rust | 2Lr67 | | cyrille saintenac | | Thatcher | Chinese spring |
| septoria | 3stb6 | | cyrille saintenac | | Chinese spring | Obelisk, courtot |
| septoria | 2stb16 | | cyrille saintenac | | | |
| GPD | 2NAM-A1 | | jacques legouis | | | |
| GPD | 1NADH-GOGAT | | jacques legouis | | | |
| chlortoluron (desherbant) | 2 | | jacques legouis | | cellule, lyrik, hendrix, sy_tolbiac | fanion, rubisko, bergamo, sy_moisson |

Perspectives: développement de nouveaux marqueurs diagnostiques

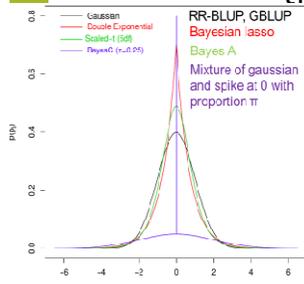
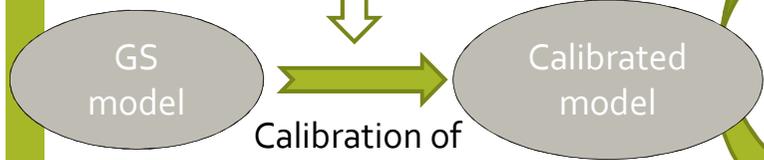
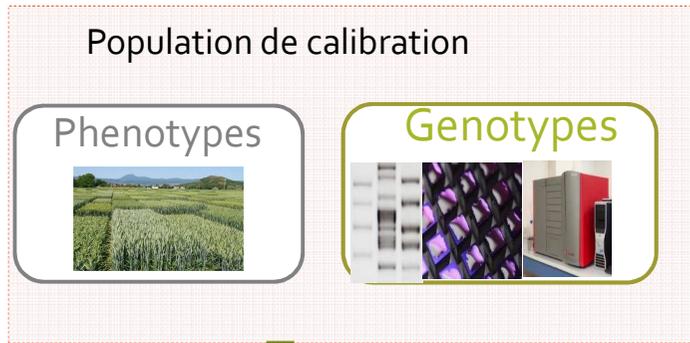
- Séquençage de 2000 gènes candidats dans le cadre de BW (450 accessions représentatives de la diversité: BWP3, 300 lignées élités)
- Séquençage d'exome whealbi
- Etudes d'association avec haplotypes sur les panels BW et Rustwatch
- Design et validation (témoins) des marqueurs
- Nouveaux caractères, jeux de données, gènes candidats?

PRÉDICTIONS GÉNOMIQUES

Caractères quantitatifs, puce 35K

Prédictions génomiques

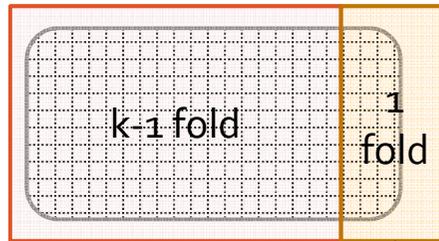
Calibration



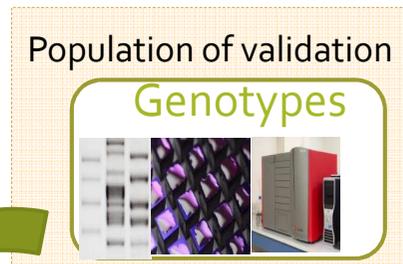
$$y = \mu + Z\beta + e$$

estimate the vector β of marker effects

Cross validation dans un panel qui a été phénotypé dans le même essai



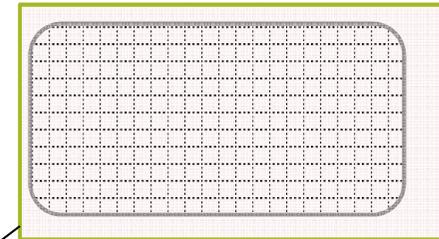
Validation



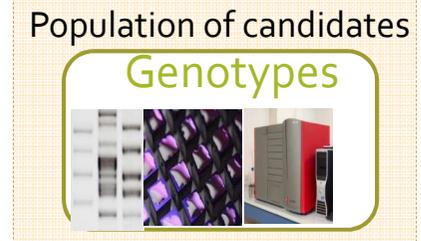
GEBVvalidation

Predictive ability = $\text{cor}(y, \text{GEBV})$

Validation sur un panel qui a été phénotypé dans un essai différent, des génotypes différents



Sélection



GEBVcandidates

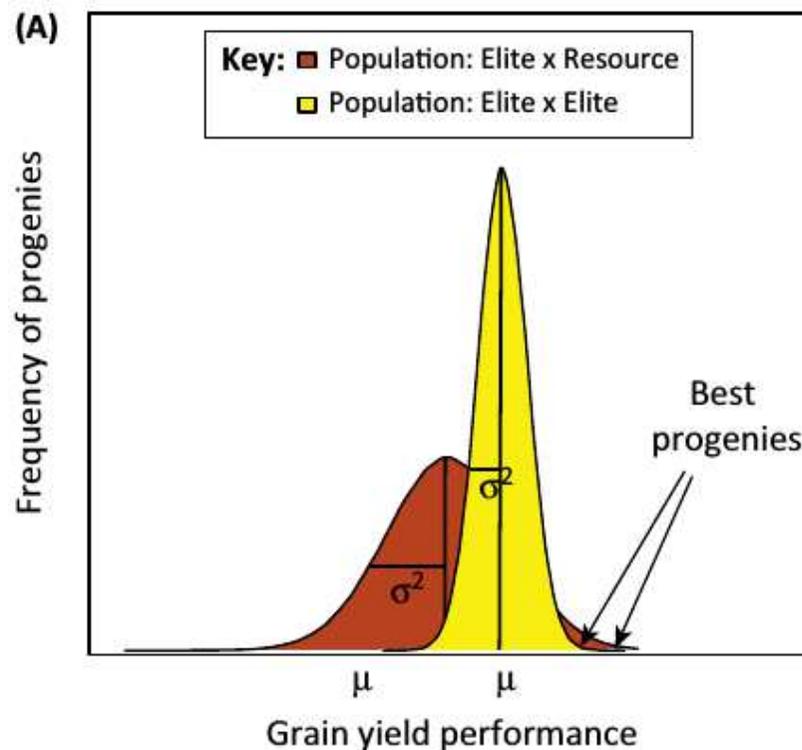
Score moléculaire
Classement des candidats

$$\text{GEBV} = Z \hat{\beta}$$

- Prédiction de valeurs individuelles
- Prédiction de la valeur des croisements élite X élite
- Prédiction de la valeur des croisements élite X exotique

Prediction de croisements

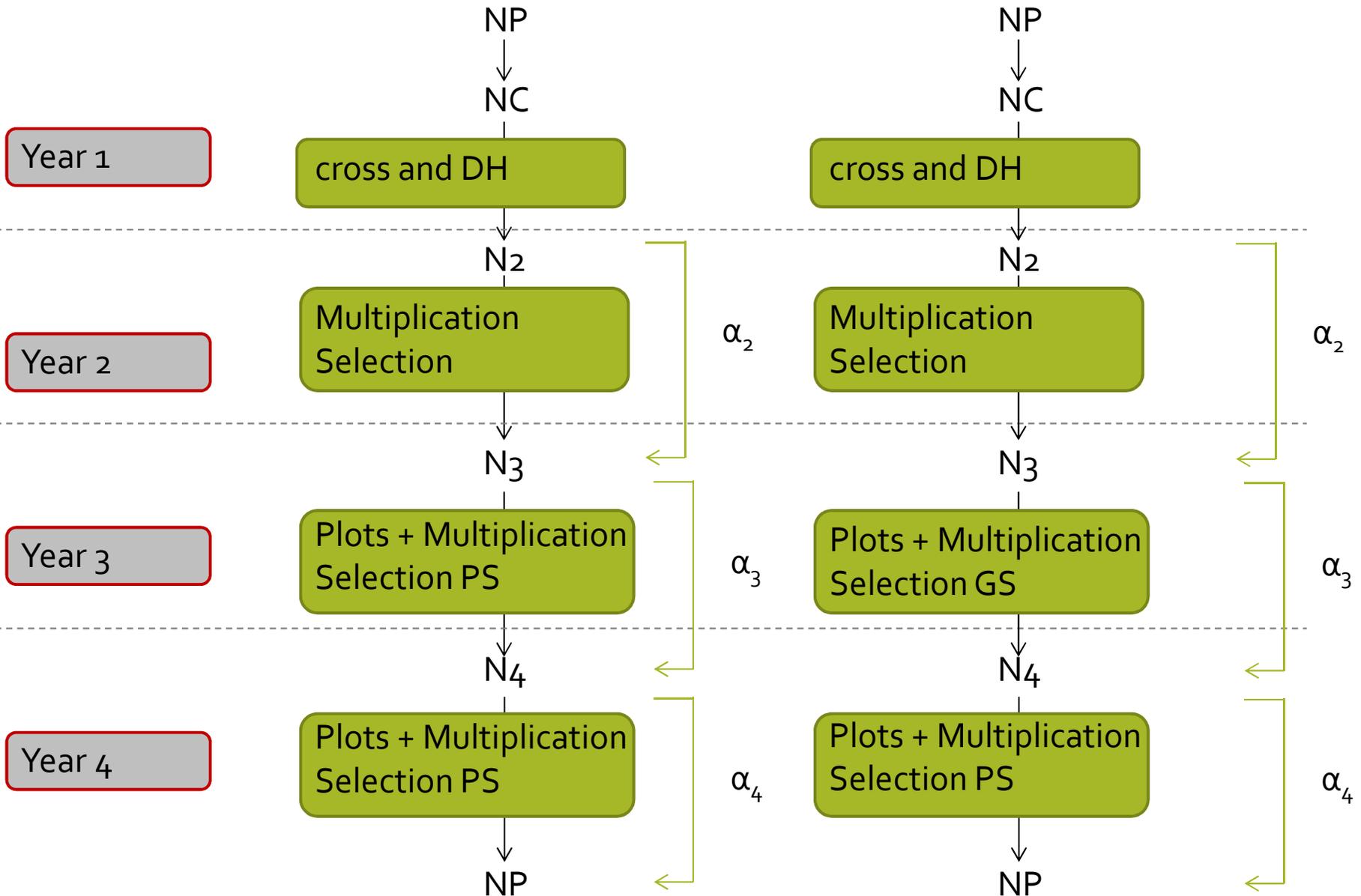
- Augmenter le gain génétiques pour des caractères complexes: probabilité d'avoir un descendant supérieur avec les effectifs qu'on a
- Contrôler et cibler la diversité utile qu'on ré-introduit



Simulation de schémas de sélection

PS scheme

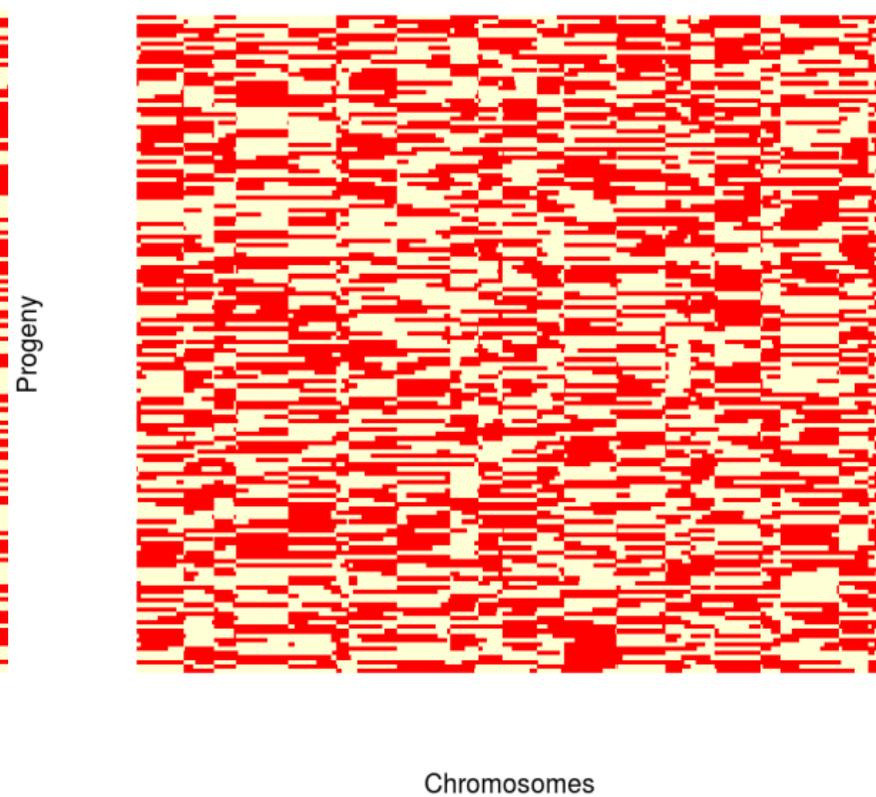
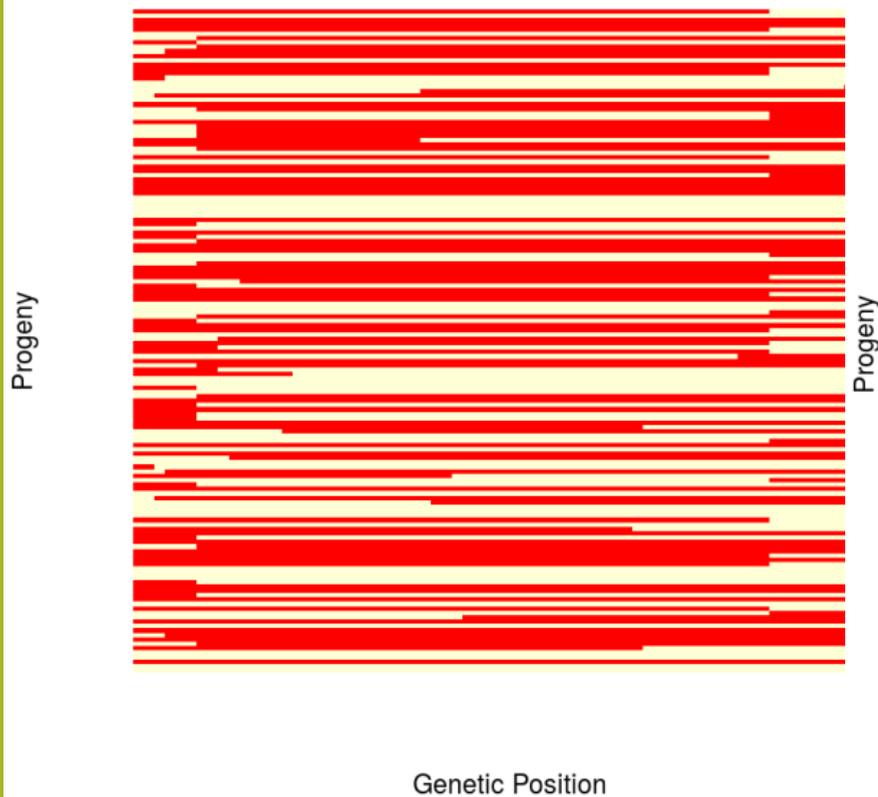
GPS scheme



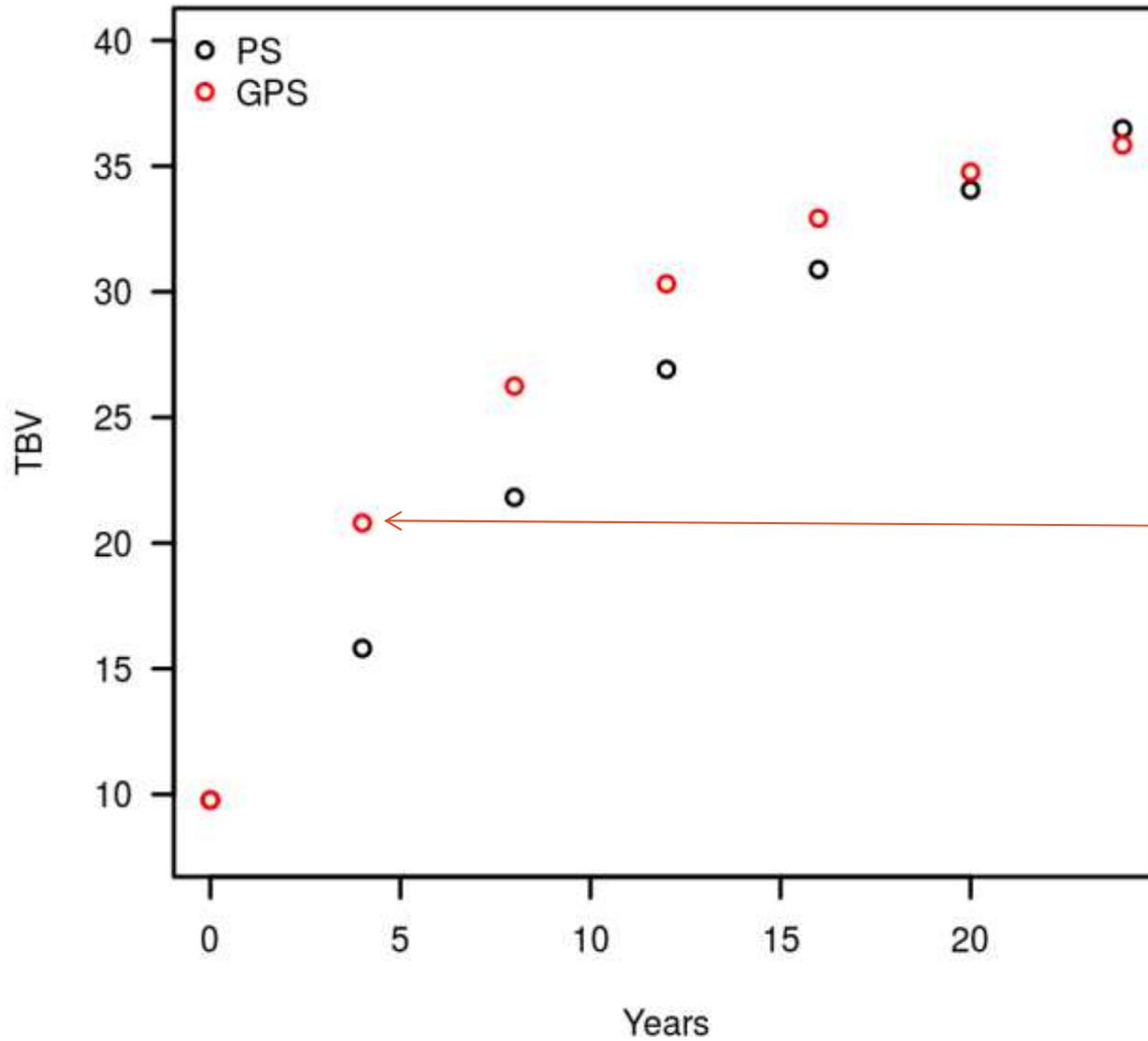
Number of plants per cross

| PS.n2 | GPS.n2 | PS.n3 | GPS.n3 | PS.n4 | GPS.n4 |
|-------|--------|-------|--------|-------|--------|
| 351 | 409 | 70 | 82 | 1.27 | 1.32 |

Simulation de croisements de lignées génotypées, production de CO et descendants in silico



lambda=0.25 h2=0.7

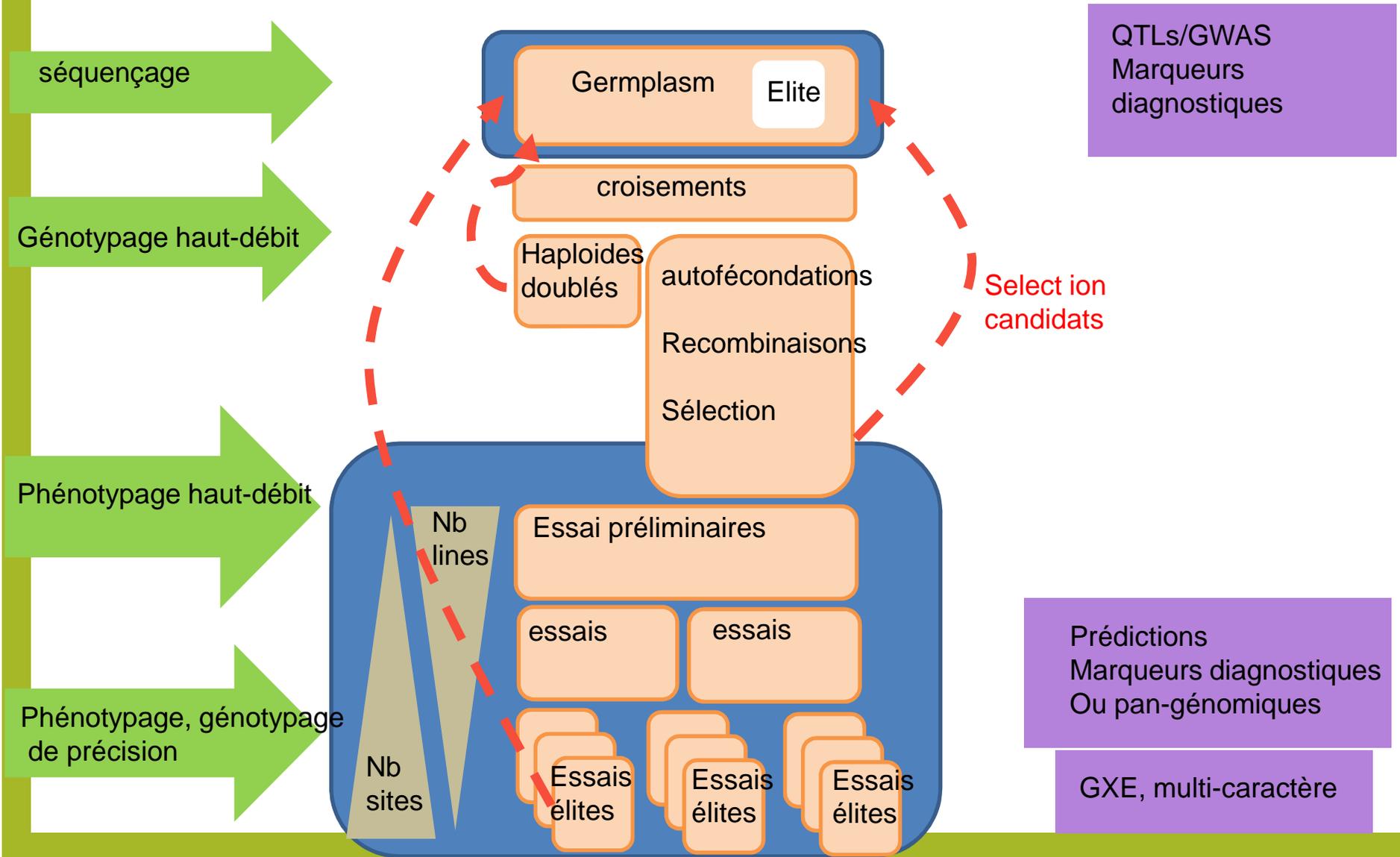


$$\alpha_{3 \text{ GPS}} = 0.27$$

$$\alpha_{4 \text{ GPS}} = 0.02$$

Prédiction des
meilleures
combinaisons
parentales

Perspectives: optimiser le schéma de sélection



MERCI

Jérôme Auzanneau
François-Xavier Oury
Emmanuel Heumez
Bernard Rolland
Jacques Legouis
Catherine Ravel
Cyrille Saintenac
François Balfourier
Etienne Paux
Gilles Charmet