



Modèles de simulation en génétique quantitative et des populations

**Apha Niango ,Hélène Muranty
Léopoldo Sanchez**

Journée CaSciModOT
13 décembre 2011 –Orléans

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Modèle de simulation en génétique

Contexte

Dans le processus d'amélioration, l'efficacité de la sélection dépend en grande partie de la précision de l'estimation des valeurs génétiques des candidats.

Plusieurs approches possibles pour identifier les méthodes de sélection les mieux adaptées aux espèces :

- Approche entièrement analytique
- Approches par simulations

Simulation en modèle allélique ou à nombre de locus fini (*s'intéresse au devenir des allèles aux différents locus*)

Modèle de simulation en génétique

Simulations Monte-Carlo : qu'est-ce que c'est ?

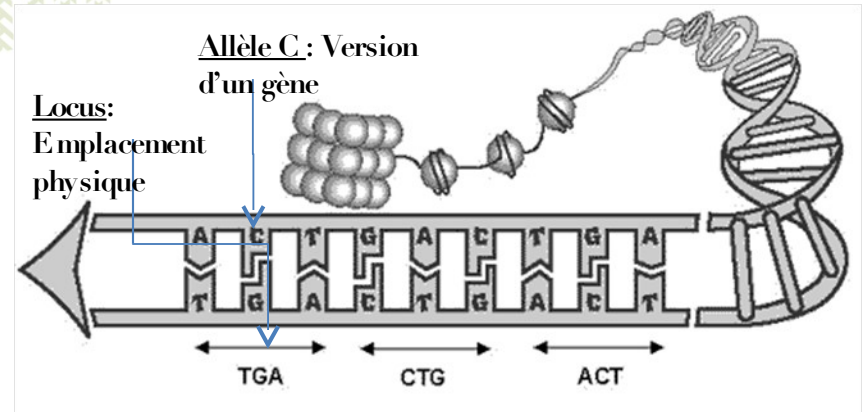
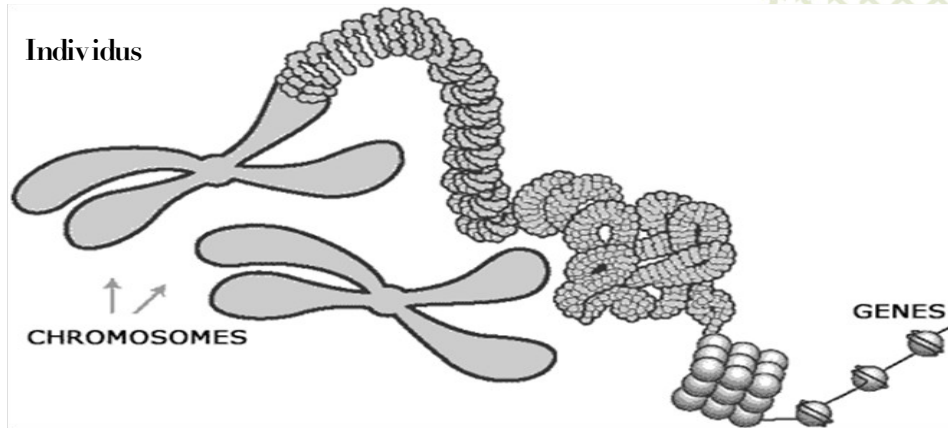
Simulation : méthode d'analyse visant à imiter un système de la "vraie vie" (quand d'autres analyses sont trop complexes ou difficile à reproduire)

Monte-Carlo : Utilisation d'expériences répétées pour évaluer une quantité ou résoudre un système déterministe (casino et jeux de hasard)

Etudier l'évolution d'un système à partir de l'échantillonnage répété des variables incertaines selon une loi de distribution.

Modèle de simulation en génétique

Modèle allélique ou à nombre de locus fini



Conceptuellement: De nombreuses hypothèses définissent les mécanismes sous-jacents

En pratique: programmation complexe et temps de calcul importants

Basé sur des gènes simulés individuellement, avec des effets génétiques donnés, dans une architecture génétique prédéfinie et organisés dans un génome virtuel.

Modèle de simulation en génétique

Echantillonnage aléatoire

L'échantillonnage/Simulation aléatoire est appliqué à différentes phases du processus simulé:

Construction de génomes virtuels à partir des gènes individuels (échantillonnage aléatoire des fréquences, effets et allèles) pour définir des individus opérationnels (prêts à être sélectionnés et croisés)

Modélisation du devenir de chaque gène (par recombinaison et ségrégation virtuelles pour obtenir de nouveaux individus opérationnels)

Modèle de simulation en génétique

Echantillonnage aléatoire (Construction de génomes virtuels)

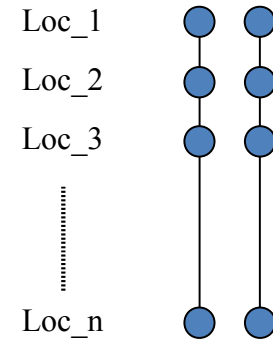
Echantillonnage des allèles

Selon des fréquences alléliques

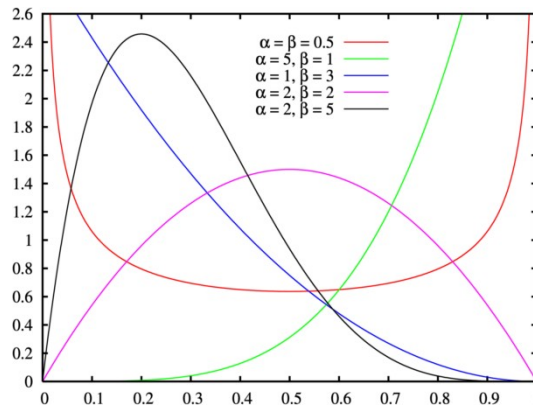
- définies en entrée
- tirées dans une distribution uniforme [a,b]

- Tirages dans d'autres distributions.
- Plus représentatif de la réalité.

Hasard



Génome
des fondateurs



Processus à modéliser ...

fréquences alléliques carte génétique

Formateurs

stratégie

Croisements

Sélection

Méiose - Gamétogénèse

Descendants

index multicaractère, niveaux indépendents
consanguinité selon pedigree, proximité selon marqueurs

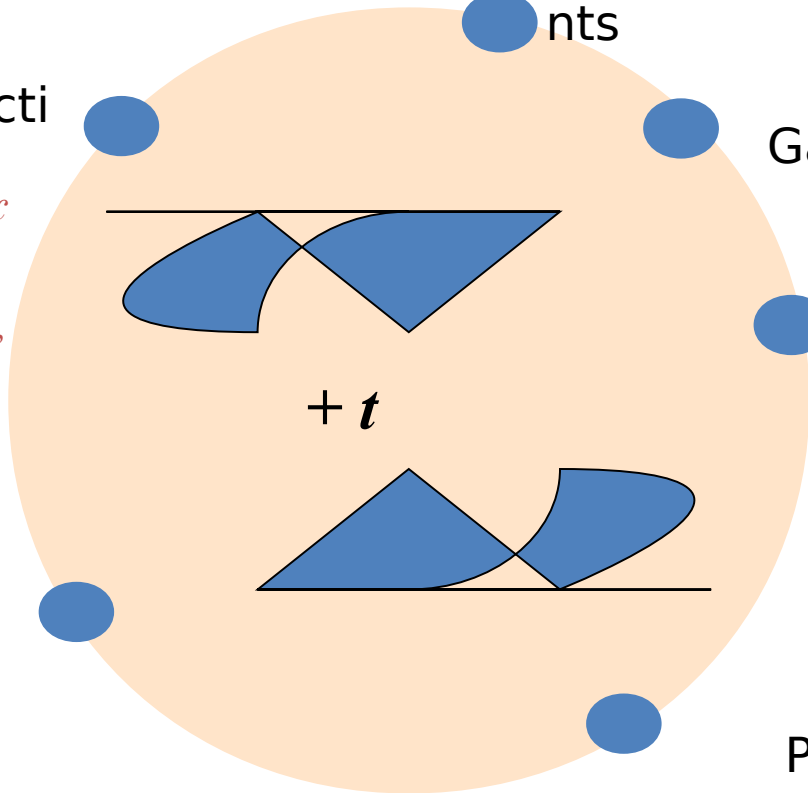
+ *t*

architecture génétique (effets génétiques, corrélations, héritabilités)

Evaluation génétique

index individu-famille, BLUP, MA-BLUP

Phénotypes/ Caractères



Processus à modéliser ...

fréquences alléliques carte génétique

Formateurs

stratégie

Croisements

Sélection

Méiose - Gamétogénèse

index multicaractère, niveaux indépendents

consanguinité selon pedigree, proximité selon marqueurs

Descendants

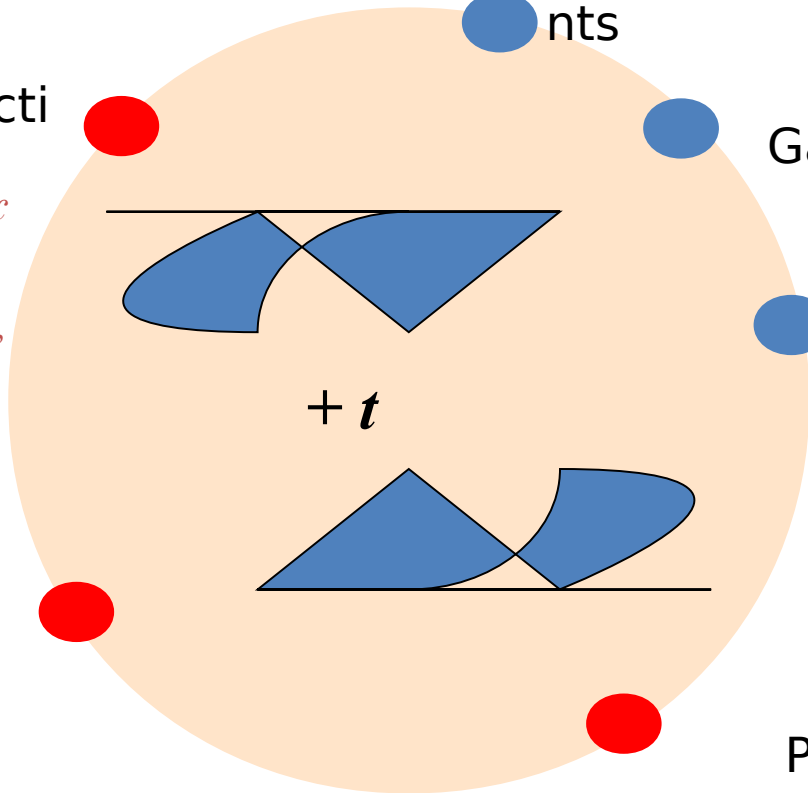
+ *t*

architecture génétique (effets génétiques, corrélations, héritabilités)

Evaluation génétique

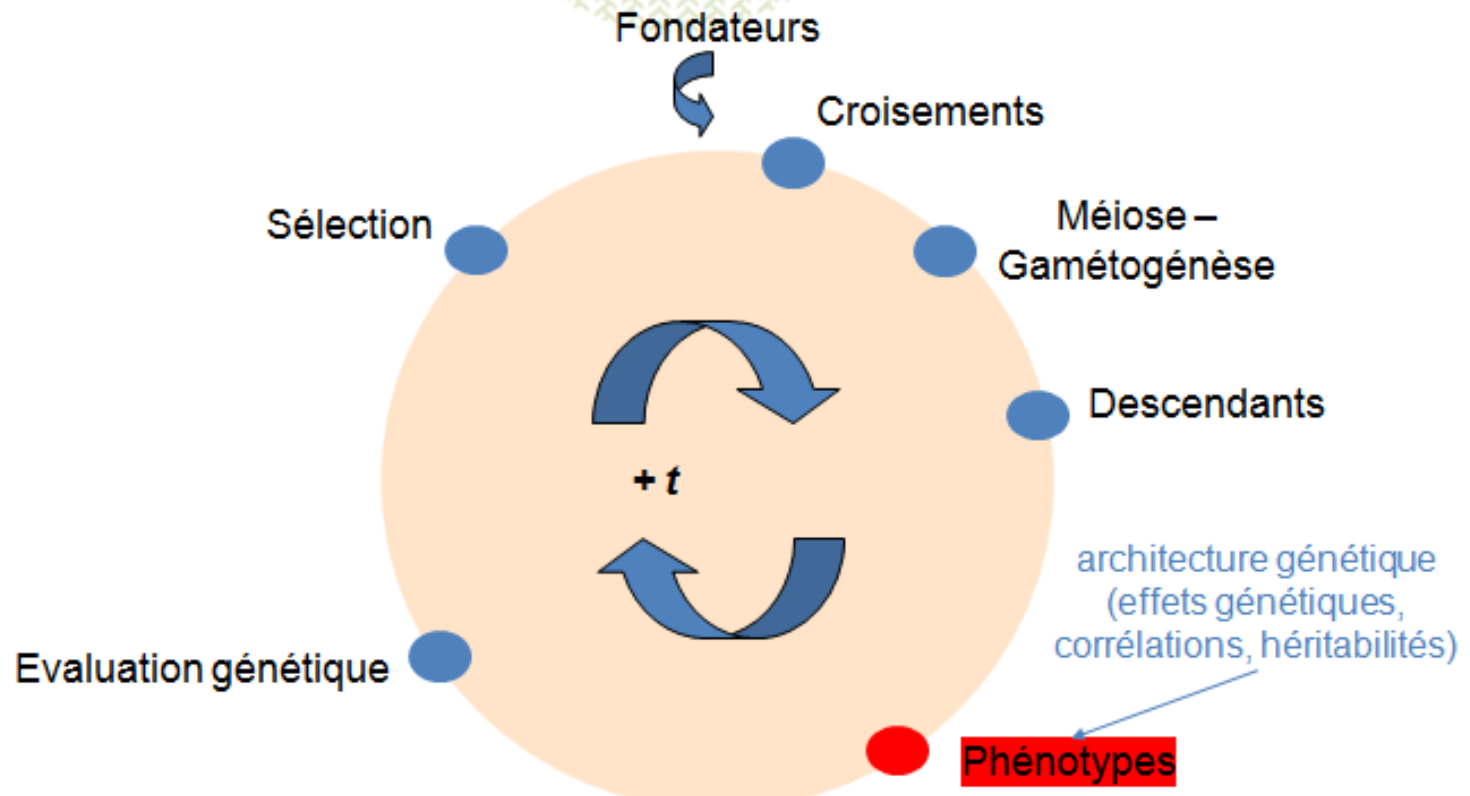
index individu-famille, BLUP, MA-BLUP

Phénotypes/ Caractères



Modèle de simulation en génétique

Type d'architecture génétique



Modèle de simulation en génétique

Type d'architecture génétique pour l'étude de la plasticité

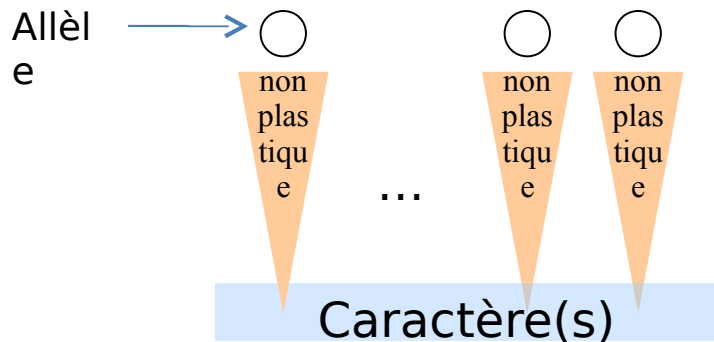
Plasticité: Capacité d'un génotype à produire différents phénotypes en fonction de l'environnement.

Le modèle de base qui explique la « genèse » d'une performance ou phénotype est sous forme linéaire:

$$P = G + E$$

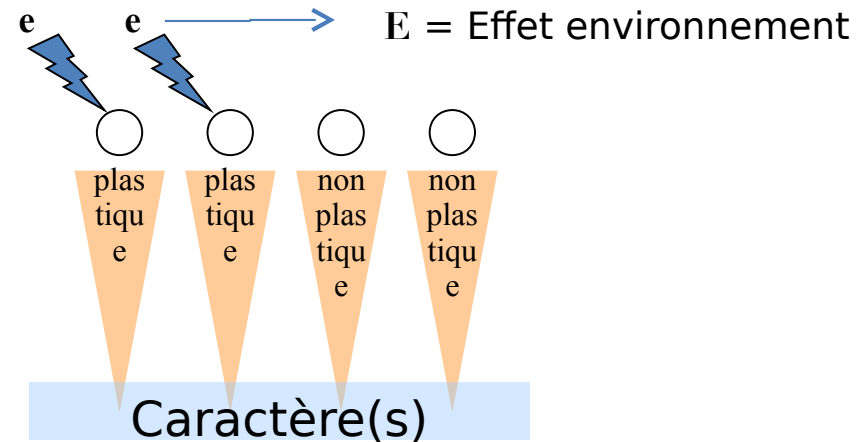
Performance *Effet du génotype* *Effet du milieu*

Réseau non structure



a_i : Effet moyens de l'allèle i

$$P = \sum(a_i) + \xi$$



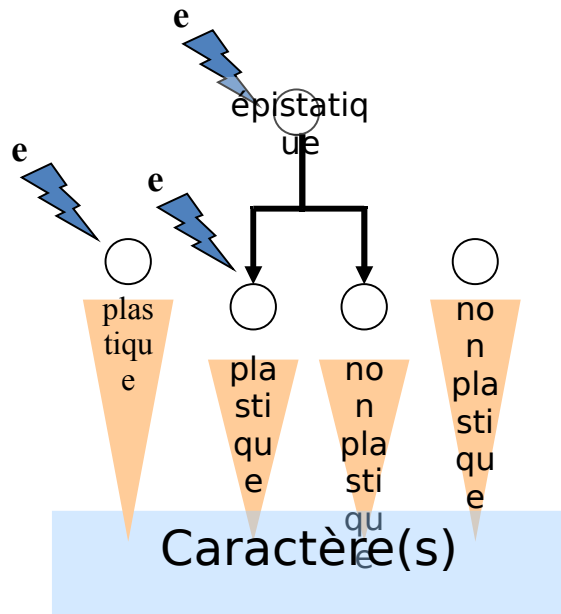
$$P = \sum(a_i) + \sum(a_i \times e) + e + \xi$$

Modèle de simulation en génétique

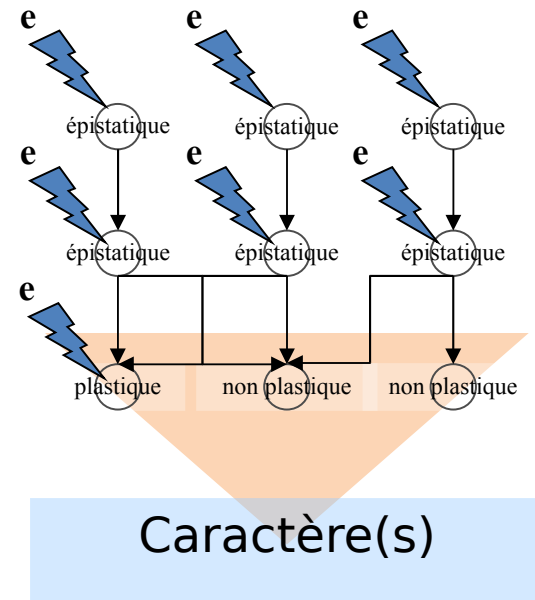
Type d'architecture génétique pour l'étude de la plasticité

Epistasie: Interaction entre gènes dans le contrôle d'un caractère.

Réseau faiblement structuré



Réseau fortement structuré



$$P = \underbrace{\sum(a_i)}_{\xi} + \sum(a_i, a_u) + \sum(a_i, e) + e +$$

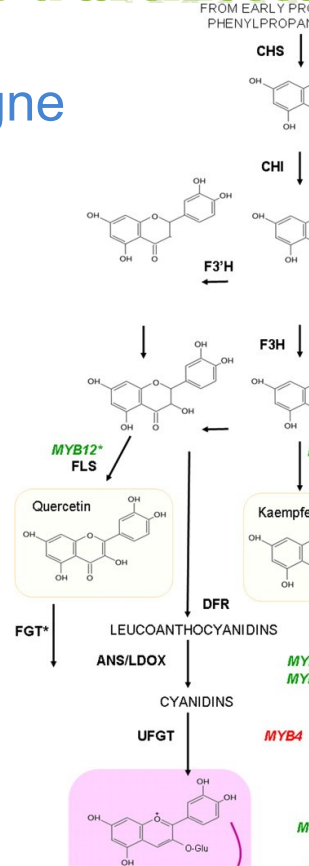
Modèle de simulation en génétique

Type d'architecture génétique complexe

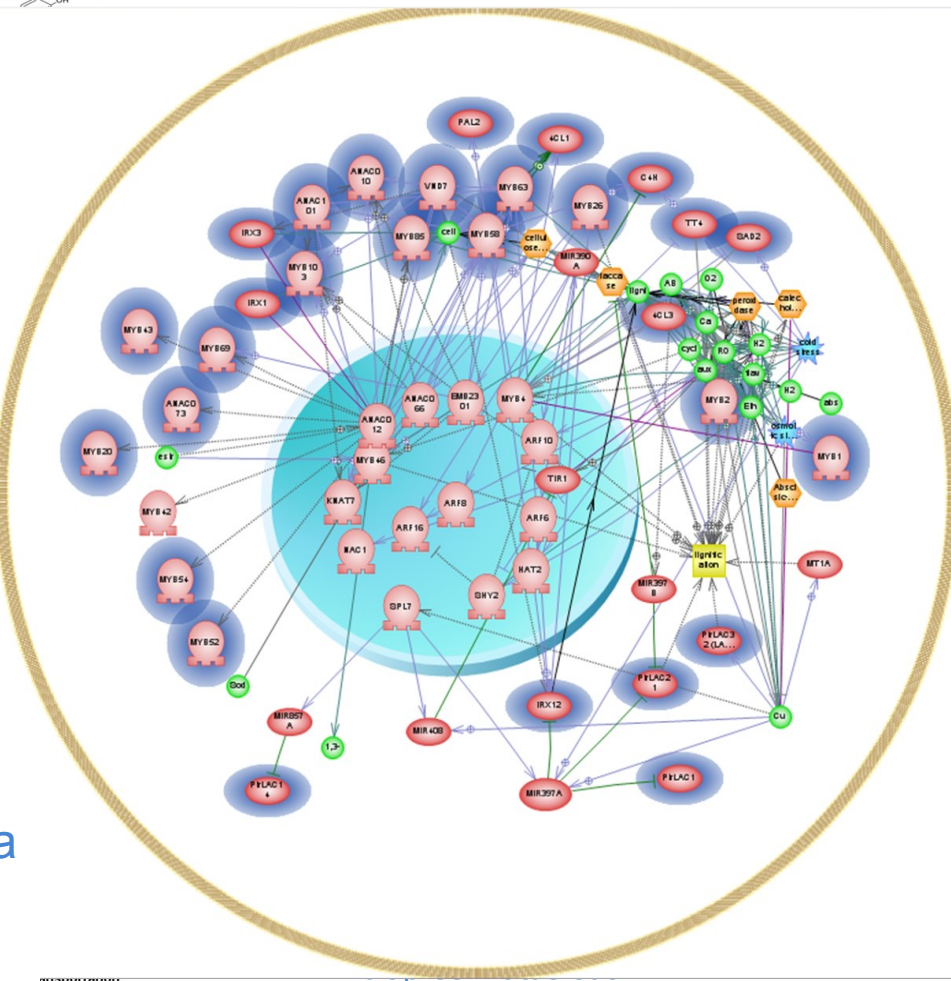
Voie métabolique des anthocyanines chez la vigne

ENZYMES

RÉGULATEUR
S



réseau de régulation de la biosynthèse des lignines d'après JC Leplé

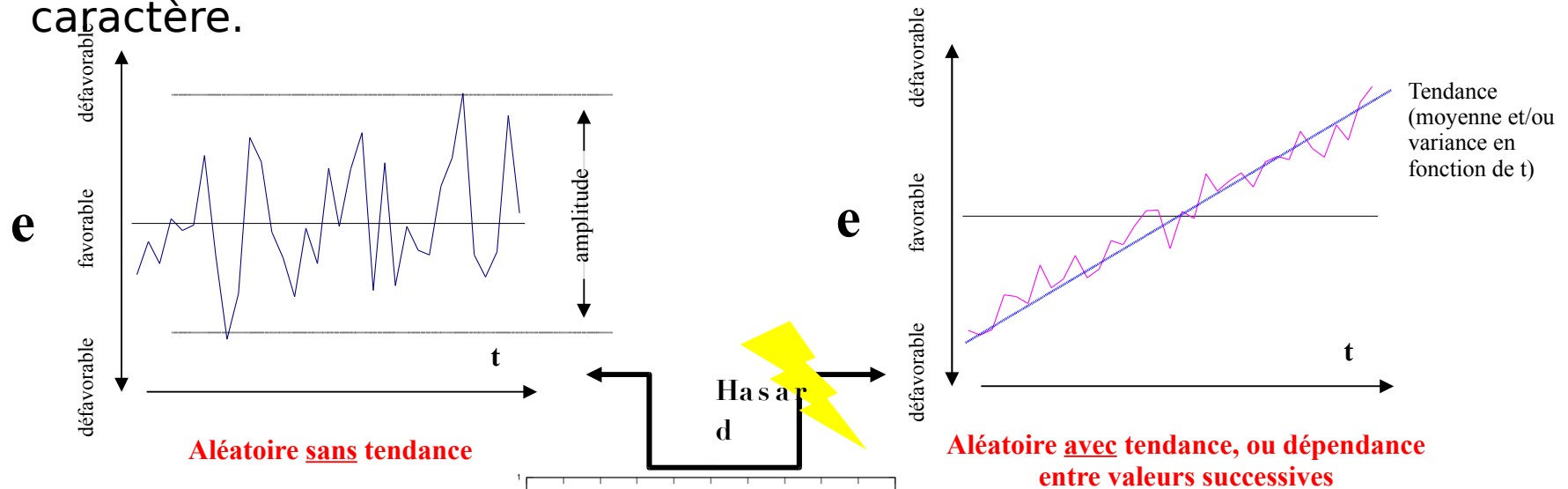


(2009)

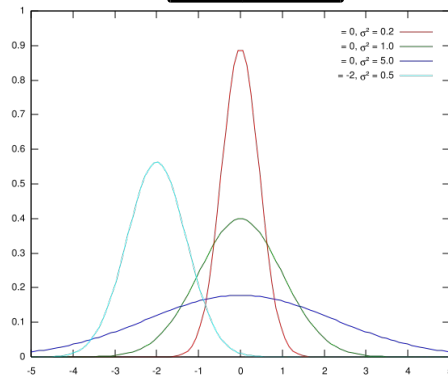
Modèle de simulation en génétique

Architecture génétique complexe pour l'étude de la plasticité

Evolution des valeurs environnementales dans le temps pour un caractère.



Tirage aléatoire des valeurs e d'une distribution normale, moyenne et variance connues

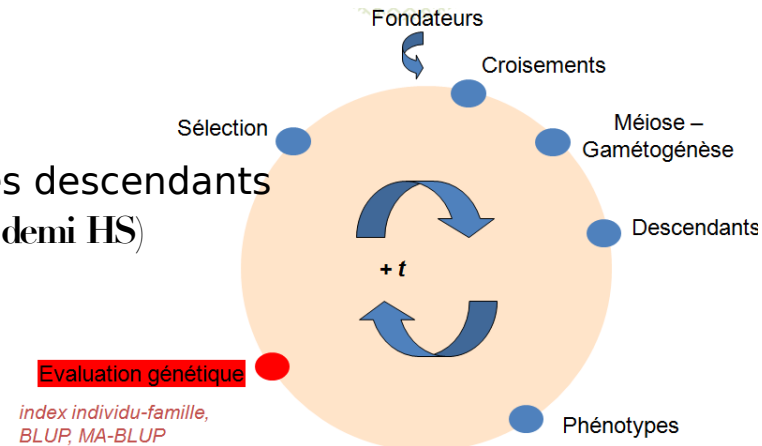


Modèle de simulation en génétique

Evaluation génétique

Un candidat à la reproduction est choisi selon certains critères. L'approche de sélection est fondée sur des index de valeur génétique construits à partir de différente(s) information(s)

- ❖ Performance(s) individuelle(s)
- ❖ Performance(s) des parents
- ❖ Performance moyenne d'un petit échantillon des descendants
- ❖ Performance moyenne des frères et sœurs (pleins FS et demi HS)



Index individu-famille

$$\text{Phénotype} \leftarrow Y = \mu + HS + FS/HS + \text{ind}/FS/HS + \varepsilon$$

- Utilise seulement la structure en familles de la dernière génération
- Peu fiable en cas d'effet d'environnement commun
- Peu précise


→ résidus


Modèle de simulation en génétique

Evaluation génétique

modèle animal "classique" : BLUP

$$Y = \mu + Zu + \varepsilon$$

effet "animal" 

$$\text{Var}(u) = A \sigma^2 a$$


matrice d'apparentement calculée à partir du pedigree (généalogies ou connaissance des filiations)

- calcul de la matrice d'apparentement
- résolution du modèle mixte par REML (maximum de vraisemblance résiduel) pour estimer la composante de la variance $\sigma^2 a$
- \hat{u} meilleur prédicteur linéaire sans biais (BLUP) de u .
- \hat{u} est la valeur génétique ou index de valeur génétique prédite

Logiciels de référence

- BLUPF90 (Misztal)
- WOMBAT (Meyer)
- ASReML

Programme de simulation allélique

Adaptation au programme

Modèle de simulation en génétique

Evaluation génétique : MA-BLUP

Modèle SAM

$$Y = \mu + Zu + Wv$$

effet + ε
"animal"

↑
effet
QTL

↑

$$\text{Var}(u) = A$$

$$\sigma^2_a$$
$$\text{Var}(v) = G$$
$$\sigma^2_{QTL}$$

matrice d'identité par descendance au QTL
calculée à partir des marqueurs (et du
pedigree)

- calcul de la matrice d'apparentement
- résolution du modèle mixte par REML
- $\mathbf{u}' + \mathbf{v}'$ est la valeur génétique ou index de valeur génétique prédite

Logiciel de référence

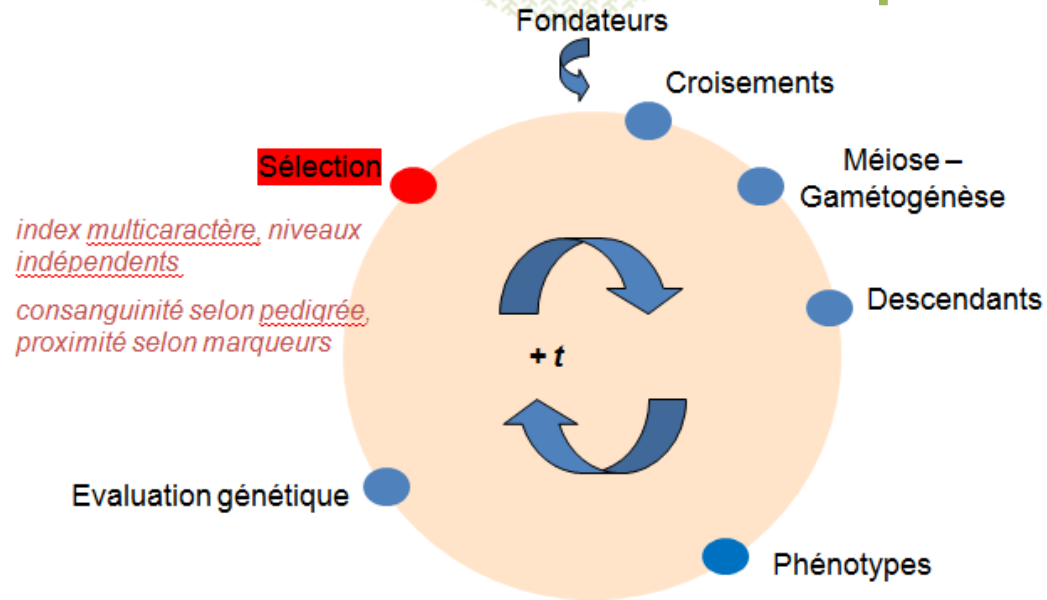
· Rtools (Pong-Wong)

Programme de simulation allélique

Adaptation au programme

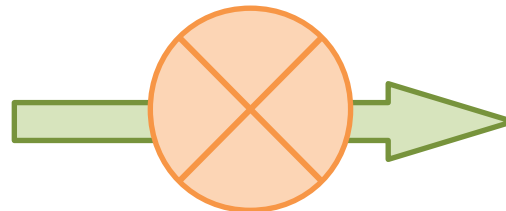
Modèle de simulation en génétique

Optimisation: Sélection en tenant compte de la diversité



approches utilisant seulement le pedigree

Liste des candidats
 - valeurs génétiques estimées g_i
 - matrice d'apparentement (a_{ij})

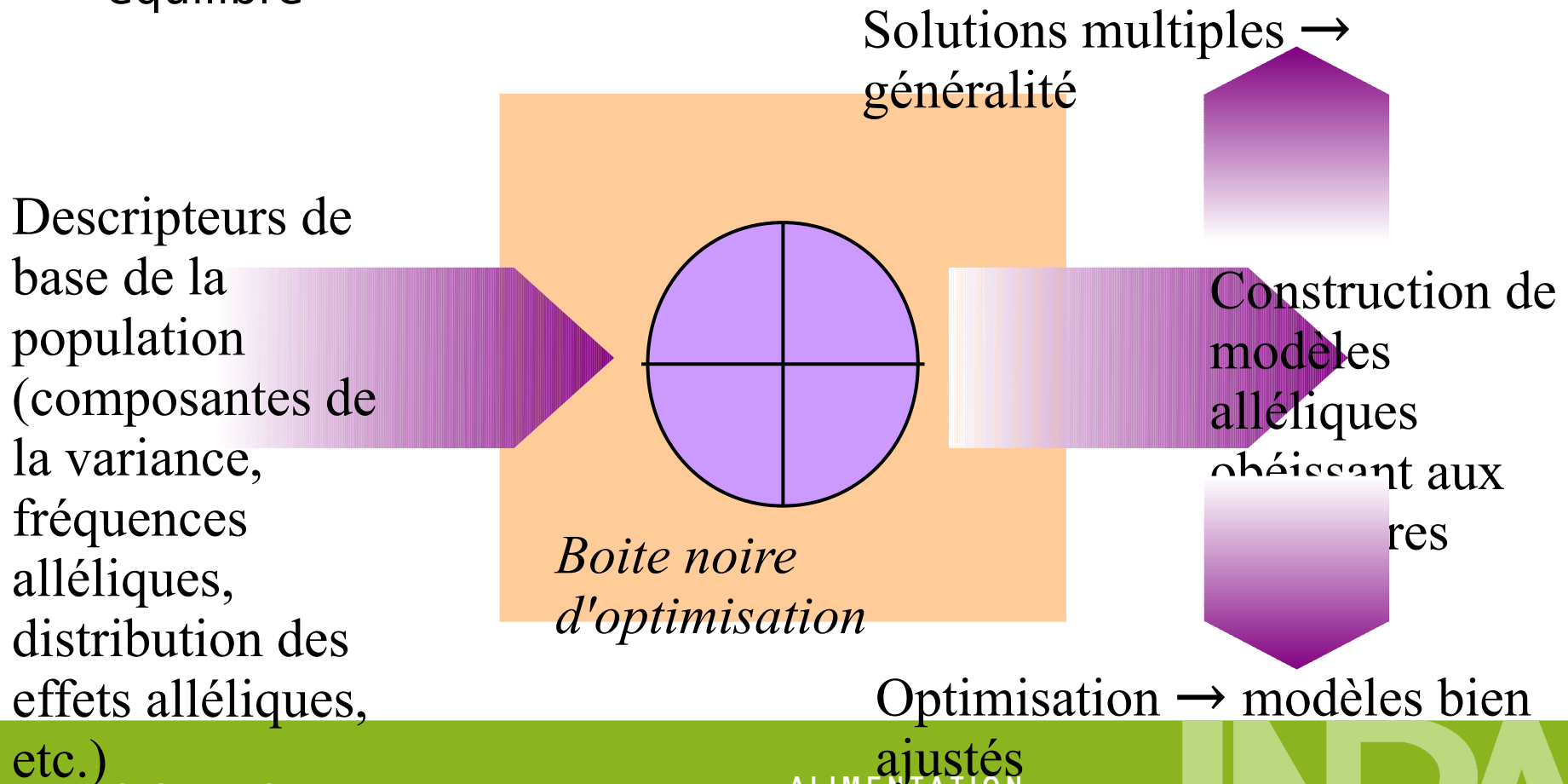


Optimisation (boîte noire)

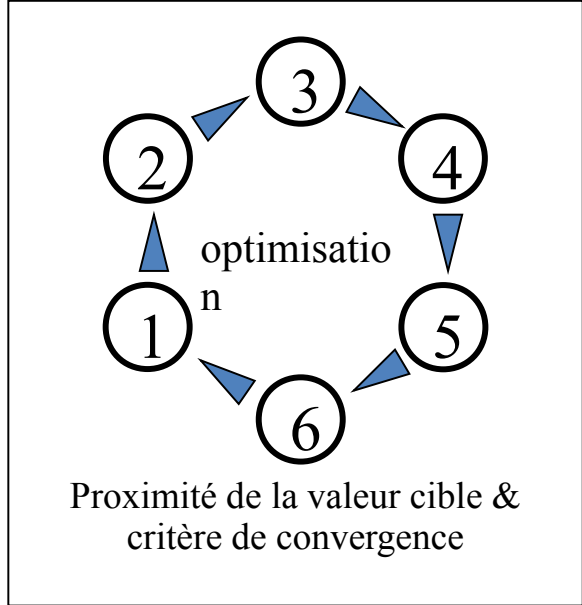
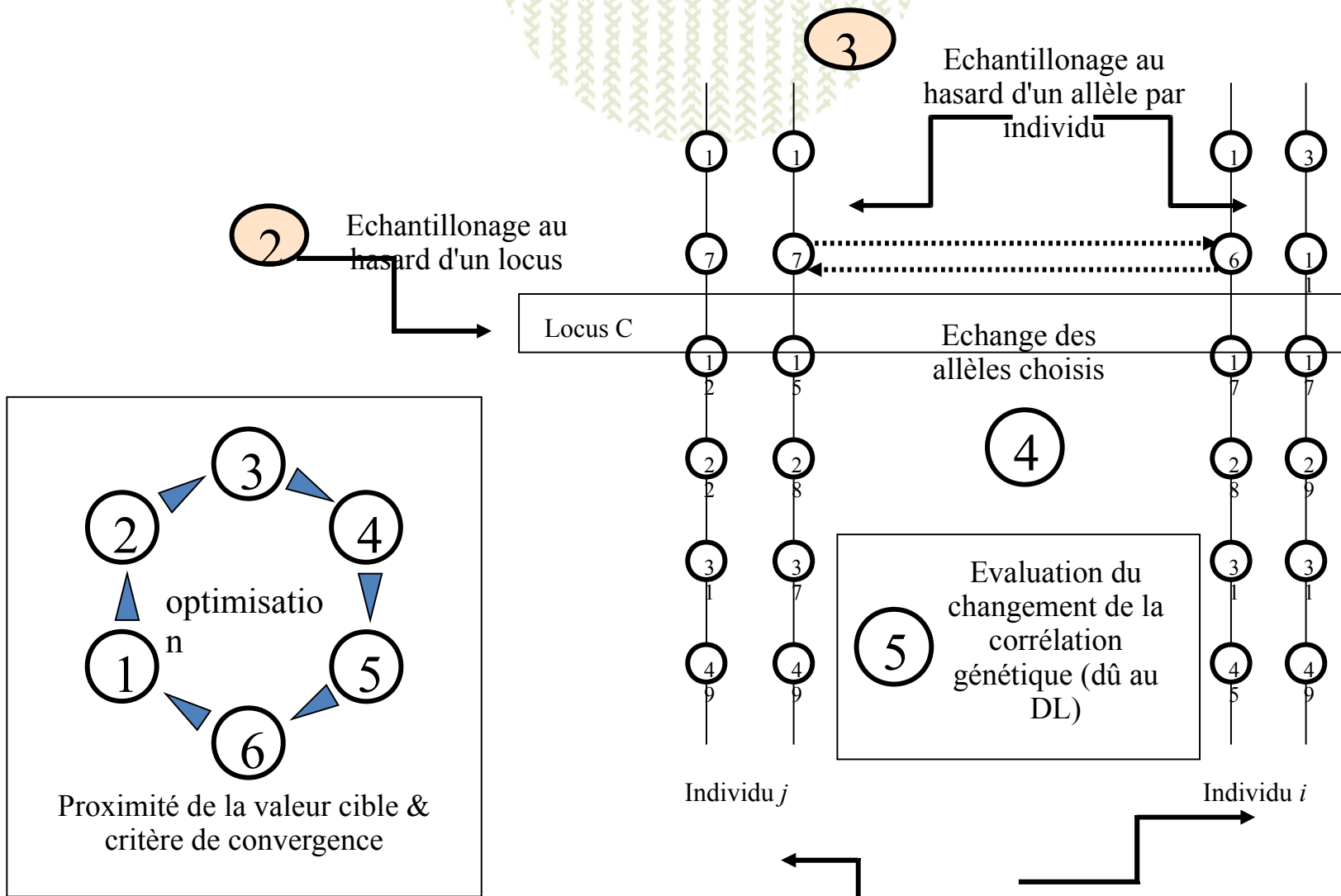
Liste d'individus sélectionnés pour maximiser $\Delta H = \sum g_i - \lambda \sum_{j,i} a_{ij}$

Modèle de simulation en génétique Optimisation

- Simulations d'un grand nombre de générations jusqu'à un équilibre



Optimisation pour obtenir une corrélation initiale donnée



1 Echantillonnage au hasard d'un couple d'individus

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



PERSPECTIVES

- **Optimisation** « ressource système (parallélisation de certaines étapes du processus) »

Grande quantité de données alléliques , répétition de la simulation allélique

Ré-optimiser a partie optimisation pour avoir un gain en temps.

- **Approche Bayésienne**

approche Chaîne de Markov par Monte Carlo pour le calcul de la matrice IBD

- **Réseau physiologiste et réseau génétique**

Passage d'un réseau physiologiste à celui génétique, Prédiction d'un phénotype connaissant du réseau



Merci de votre attention

Journée CaSciModOT
13 décembre 2011 –Orléans

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA