

# Etude d'association dans des variétés synthétiques : cas du gène GAI et de la croissance foliaire chez le ray-grass anglais

Philippe Barre

D'après la thèse de Jérôme Auzanneau

Centre INRA Poitou-Charentes

Unité de Recherche Pluridisciplinaire Prairies et Plantes  
Fourragères

86600 Lusignan

Directeur de thèse : Christian Huyghe

# Plan

- Contexte général
  - Les prairies
  - Le déséquilibre de liaison
  - Les études d'associations
- Partie 1 : Evaluer dans des variétés synthétiques les caractéristiques nécessaires pour faire des études d'association
- Partie 2 : Une étude d'association dans une variété synthétique
- Conclusion

# Contexte général

Agriculture durable

Nourrir les herbivores



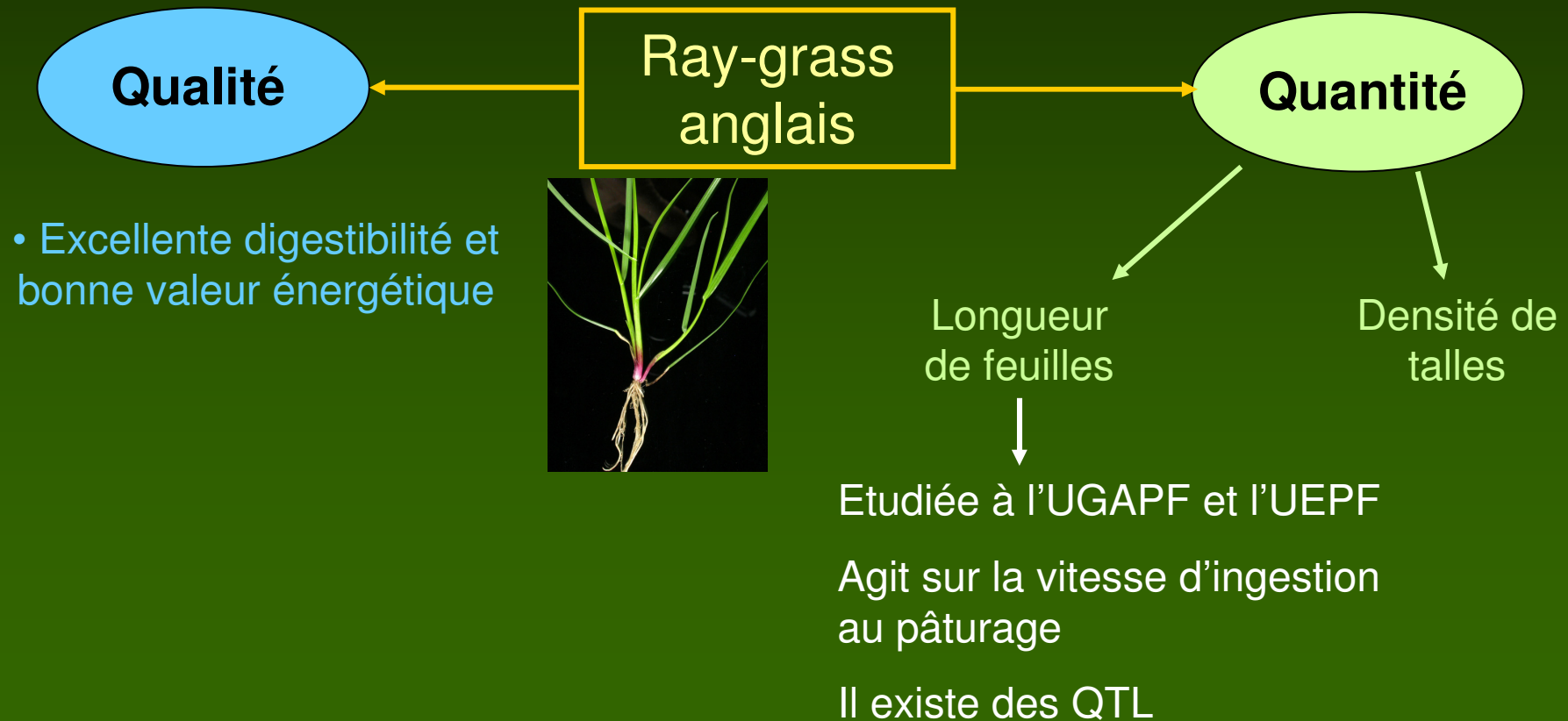
**Prairies semées :**  
pâturées ou  
fauchées



- Source d'énergie et de protéine la moins onéreuse
  - Impacts environnementaux positifs
- ⇒ sous réserve de pratiques raisonnables

Le ray-grass anglais est la graminée la plus utilisée dans les prairies semées en régions tempérées humides

# Deux facteurs importants pour nourrir les herbivores



# Deux méthodes pour identifier les gènes ou les régions du génome qui contrôlent les caractères

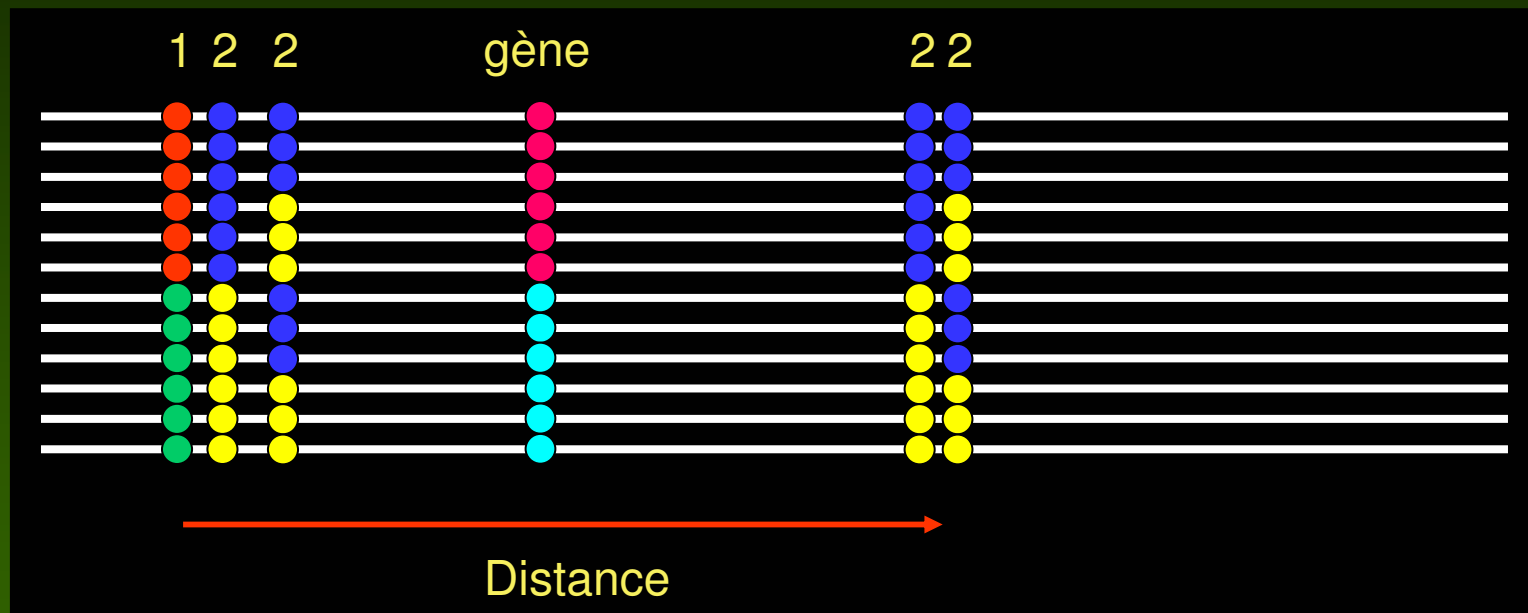
- Recherche de QTL à partir de populations dérivées d'hybrides F1
  - Bases génétiques étroites (deux parents)
  - Peu de recombinaisons
  - Création des populations parfois difficile
- Etude d'association Phénotype/Génotype sur des populations naturelles
  - Fonction du déséquilibre de liaison (DL) de la population étudiée
  - Populations parfois structurées et détection d'associations dues uniquement à la structuration



➔ Des populations idéales seraient des populations multi-parentales et non structurées

# Le Déséquilibre de liaison (DL)

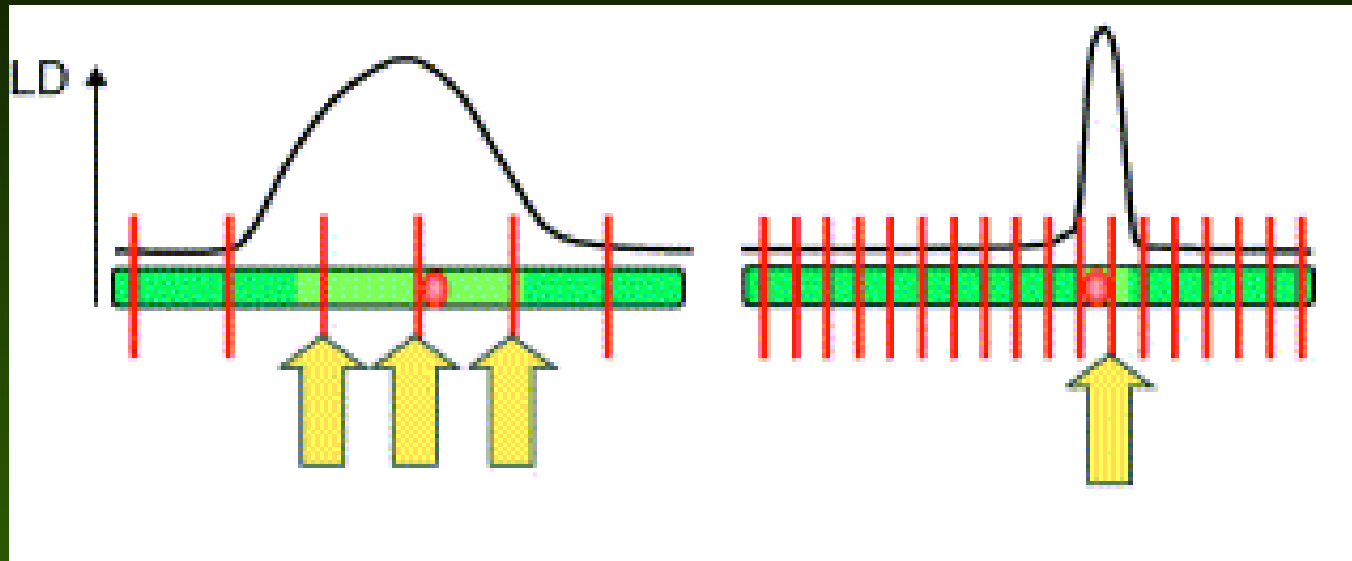
Le déséquilibre de liaison est l'association non aléatoire des allèles à deux locus



Déséquilibre complet :  $|D|=0.25$ ,  $|D'|=1.0$ ,  $r^2=1.0$

Pas de déséquilibre :  $|D|=0$ ,  $|D'|=0$ ,  $r^2=0$

# Deux approches pour les études d'associations



(Rafalski, 2002)

DL fort « Genome scan »

- exhaustive sur l'ensemble du génome
- peu précise

DL faible : « gène candidat »

- précis
- sur une région limitée

# Facteurs faisant varier le DL

(Flint-Garcia et al, 2003 ; Gupta et al, 2005)

- Taux de recombinaison et nombre de générations DL ↘
- Mode de reproduction Autogame ↗ Allogame ↘
- Histoire démographique de la population  
Goulet d'étranglement DL ↗
- Sélection DL ↗
- Taux de mutations DL ↘
- Structuration DL ↗

Pour les études d'association, on considère qu'un  $r^2 > 1/3$  entre le marqueur et le gène est suffisant pour pouvoir détecter une association (Ardlie et al., 2002)



# Prise en compte de la structuration dans les études d'association

- Evaluation de la structuration à partir de marqueurs « neutres » répartis sur l'ensemble du génome (exemple: Structure, Pritchard et al. 2000)
- Division de la population en groupes identifiés précédemment
- Recherche d'associations à l'intérieur de chaque groupe
- Limites :
  - taille des groupes
  - Variabilité intra-groupe pour les caractères étudiés



Chez le ray-grass anglais comme chez de nombreuses espèces fourragères et à gazon  
⇒ variétés synthétiques

X constituants

$4 < X < 300$  (ou plus)



3 ou 4 générations de « panmixie »

Variété synthétique  
commercialisée

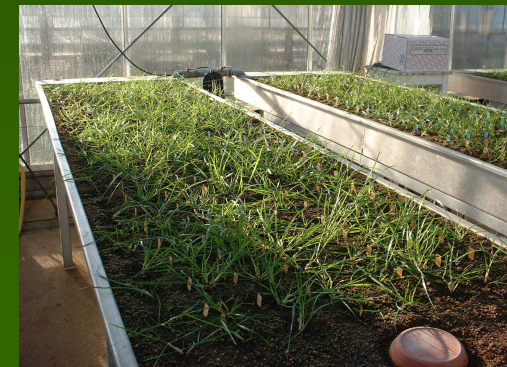
# Les variétés synthétiques semblent intéressantes pour les études d'association

- Pas de structuration
- Pas de DL entre loci non liés physiquement
- Base génétique +/- large
  - Base + large : on attend un DL qui décroît rapidement → **“gène candidat”**
  - Base + étroite : on attend un DL qui décroît lentement mais uniquement dû à une liaison physique → **“genome scan”**



## Objectif

Tester si les variétés synthétiques sont pertinentes pour identifier les gènes ou les régions du génome responsables des variations des caractères quantitatifs chez le ray-grass anglais

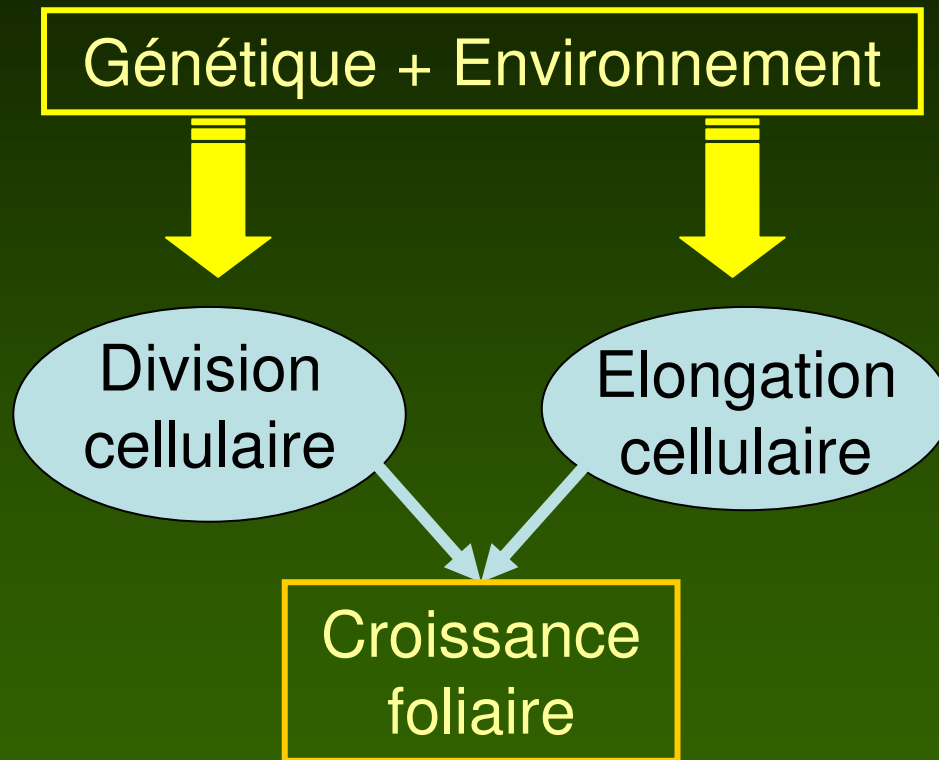


# Stratégie

- Evaluer dans des variétés synthétiques les caractéristiques nécessaires pour faire des études d'association
- Réaliser une étude d'association dans une variétés à base génétique large (Herbie) sur l'exemple d'un gène candidat Gibberelic Acid Insensitive (GAI) pour la croissance foliaire



# Les gènes candidats de la croissance foliaire



Tous les gènes qui interviennent dans :

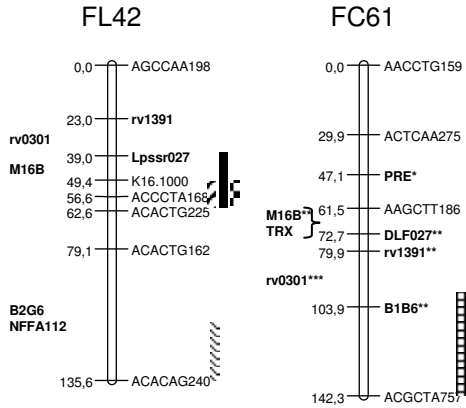
- La perception du signal
- L'intégration du signal
- La biosynthèse des hormones végétales
- Le cycle de division cellulaire
- Disponibilité en C et N

Conclusion : Il existe de nombreux gènes candidats pour la croissance foliaire

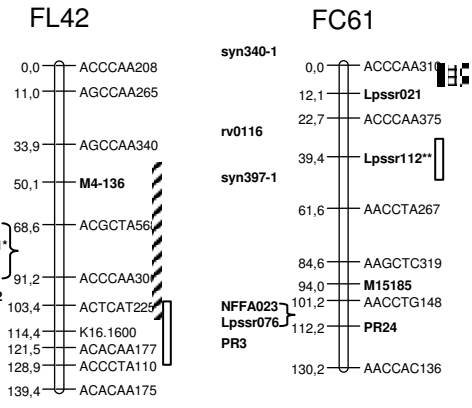
# GAI candidat pour la croissance foliaire chez le ray-grass anglais

- Impliqué dans la croissance des tiges chez de nombreuses graminées
- Chez le ray-grass anglais, croissance des tiges et croissance des feuilles sont liées
- Il existe un QTL de longueur de feuilles sur LG4 qui explique 5 à 10 % de la variabilité
- GAI est cartographié sur LG4

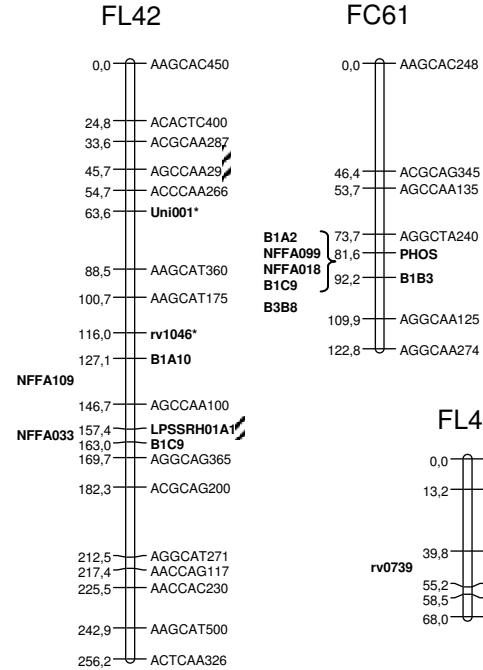
**LG1**



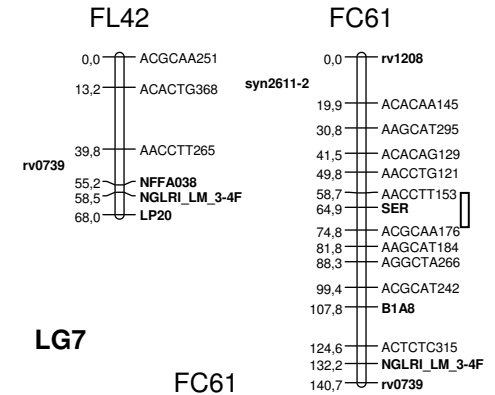
**LG2**



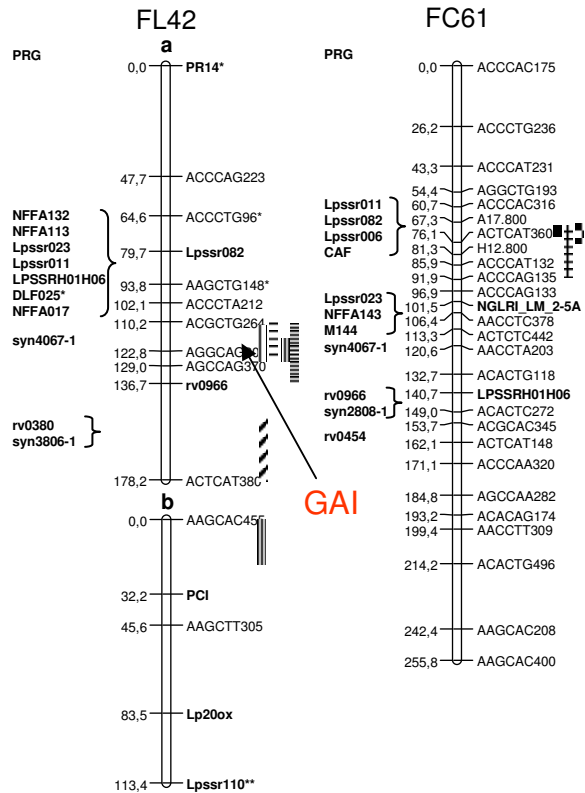
**LG3**



**LG6**



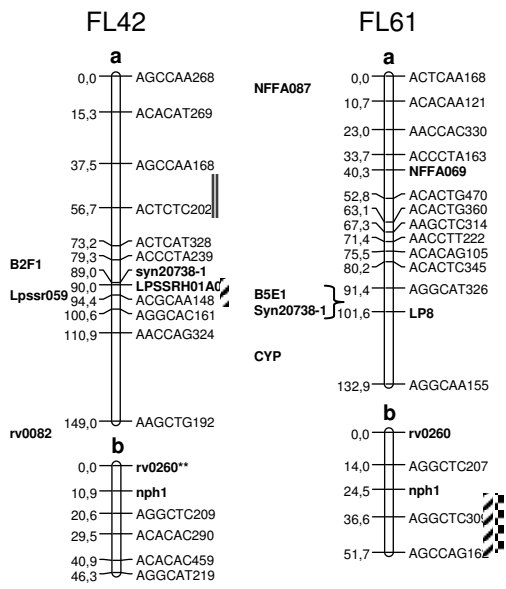
**LG4**



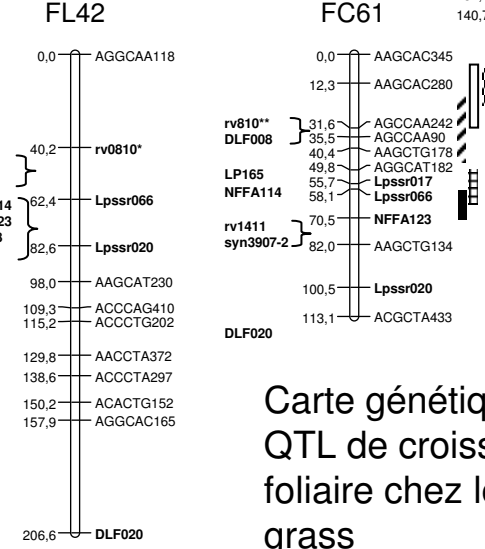
**Legend of QTLs**

- Heading00
- Leaf01
- LaminaOct99
- Lamina01
- LaminaApr00
- Sheath01
- H080300
- LER01
- H140400
- LED01

**LG5**



**LG7**



Carte génétique et QTL de croissance foliaire chez le ray-grass



# Stratégie

- Evaluer dans des variétés synthétiques les caractéristiques nécessaires pour faire des études d'association
- Réaliser une étude d'association dans une variétés à base génétique large (Herbie) sur l'exemple d'un gène candidat Gibberelic Acid Insensitive (GAI) pour la croissance foliaire



# Partie 1 : Evaluer dans des variétés synthétiques les caractéristiques nécessaires pour faire des études d'associations

Etude de 3 variétés à base génétique contrastée (troisième génération de panmixie)

- variabilité génétique et équilibre HW
- vérifier la non structuration
- déterminer le patron de décroissance du DL pour choisir le type d'approche (région de GAI)

# Matériel végétal

- 3 variétés commerciales de ray-grass anglais :
  - Aberavon : base génétique étroite (6 constituants apparentés)
  - Herbie : base génétique large (336 constituants)
  - Brest : intermédiaire (10 constituants non apparentés)
- 1 Core collection comme témoin de la variabilité existante au sein de l'espèce

# Aberavon

7 parents

1 X 2    3 X 4    5 X 6    5 X 7

8 générations

7 cycles de sélection

Passage par 2 ou 3  
femelles seulement sur  
les 4 dernières  
générations

Sélection individuelle : démarrage  
précoce, floraison tardive, résistance  
aux rouilles

Sélection sur famille : rendement,  
couverture du sol, digestibilité *in  
vitro*

6 plantes  
sélectionnées

4 générations  
de panmixie

Variété commerciale

# Herbie

4 variétés

(origine hollandaise)

Trani

Hora

Perma

Vigor

$4 \times 84 = 336$   
individus













4 générations  
de panmixie

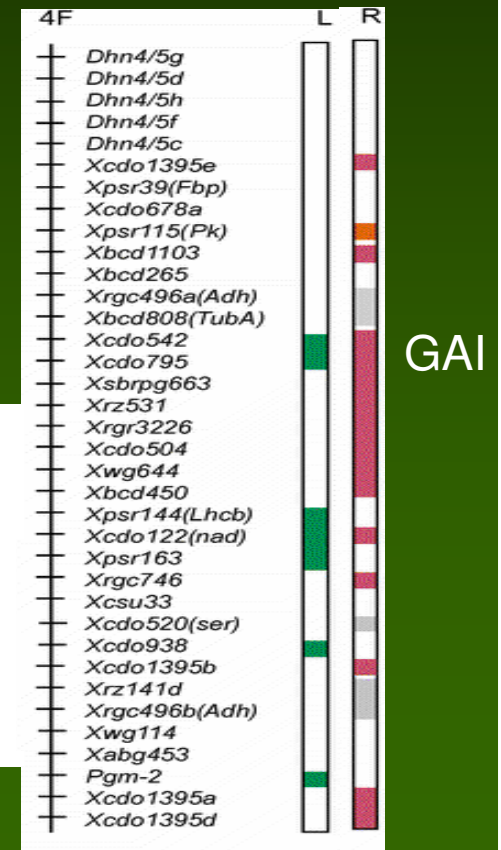
Variété commerciale

# Marqueurs moléculaires

- 6 SSR (Simple Sequence Repeat) : 3 sur LG 4, 1 sur LG 1, LG3 et LG7
  - Etude structuration, variabilité et équilibre HW
- 6 STS (Sequence Tagged Site) autour de GAI
  - Etude du DL, variabilité et équilibre HW

Code de couleur pour les chromosomes

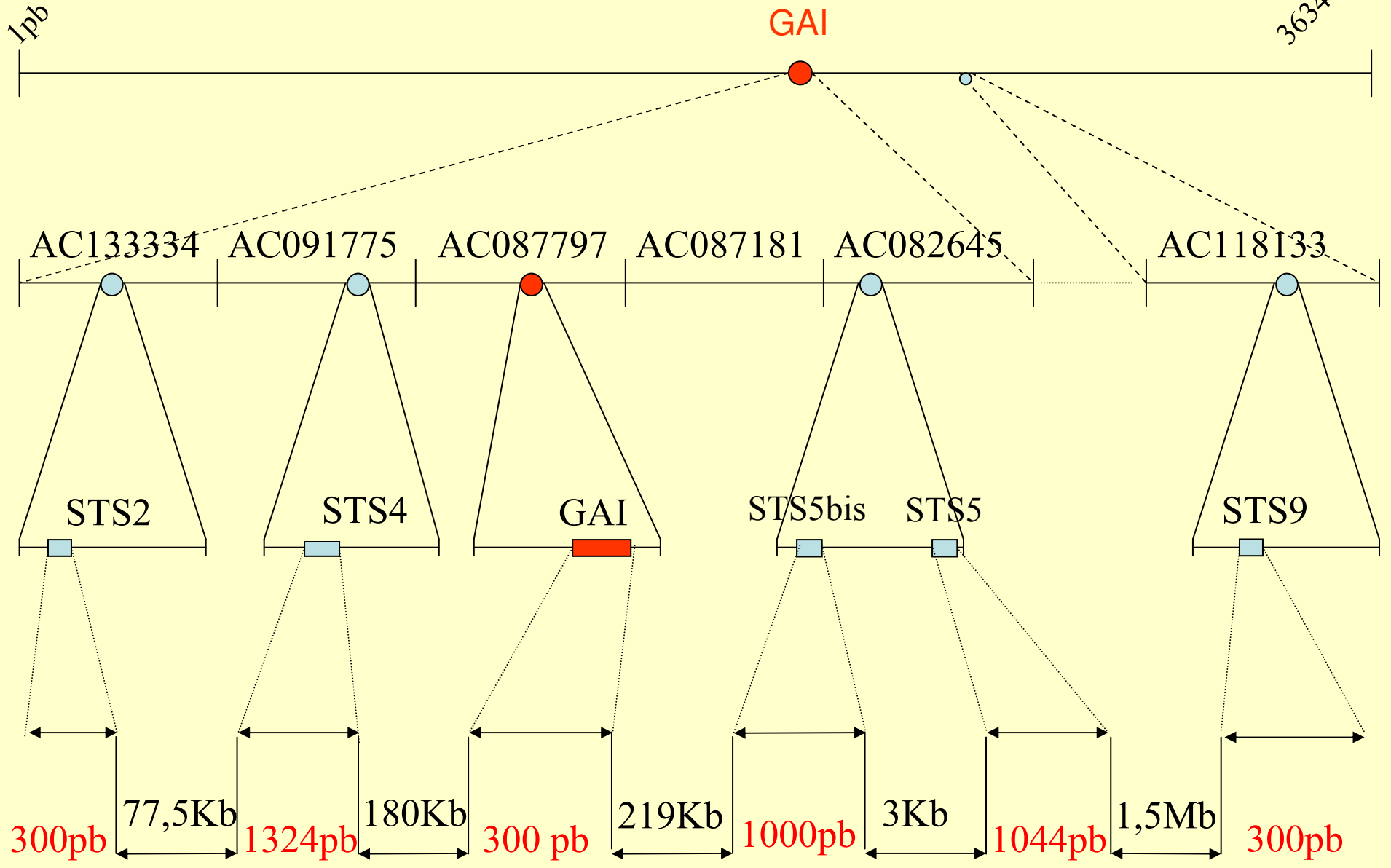
 1, A	 7, G
 2, B	 8, H
 3, C	 9, I
 4, D	 10, J
 5, E	 11
 6, F	 12



Alm et al, 2004

Pseudo molécule de riz : LG 3 (TIGR 30/12/2004) ⇒ LG 4 RGA

36347804pb

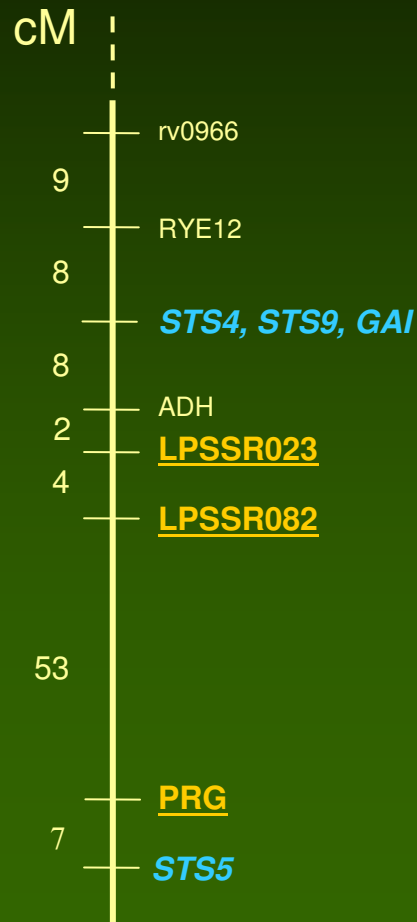


● STS

Distance chez le riz / Longueur chez le ray-grass anglais

# Distance entre marqueurs

## LG 4 ray-grass anglais






STS4, STS9 et GAI : utilisation des distances physiques (pb) observées chez le riz

STS5 et SSR : utilisation des distances génétiques (cM) observées chez le ray-grass anglais et conversion en distance physique (1 cM = 3,5 Mb)

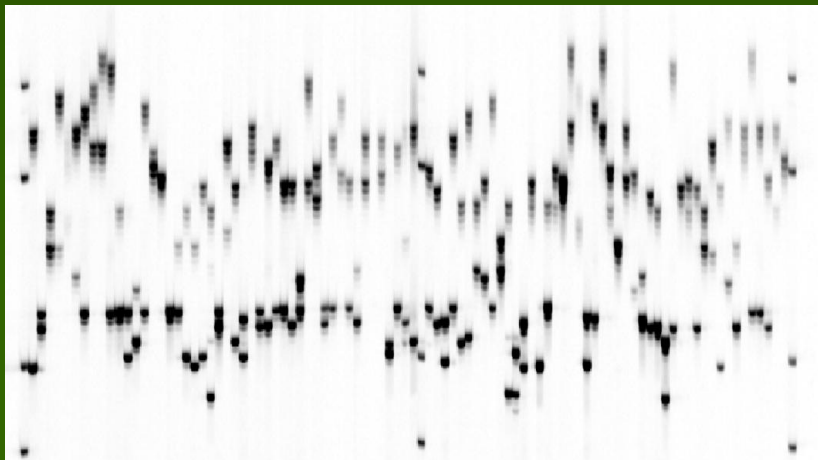
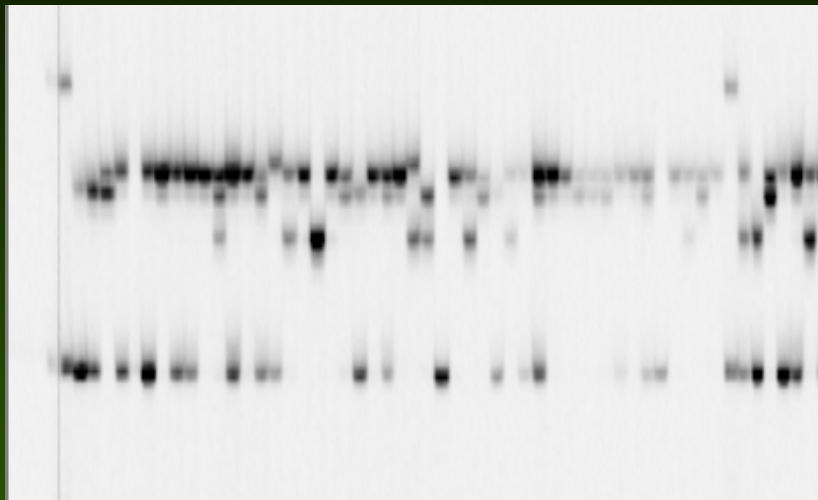


## Trois types de données

- Données de marqueurs SSR  $\Rightarrow$  génotype 
- Données de séquence de clones (STS4 et STS5)  $\Rightarrow$  haplotype 
- Données de séquence de produits PCR (~~STS2~~, STS4, STS5, ~~STS5bis~~, STS9 et GAI)  $\Rightarrow$  génotype 

Ceci pour 47 individus par population

# Données SSR



Microsoft Excel - SSR\_util\_48.xls

Eichier Edition Affichage Insertion Format Outils Données Fenêtre ?

100% Arial

Répondre en incluant des modifications... Terminer la révision...

	A	B	C	D	E	F	G	H
1			LPSSR082	B1B6	B1C9	LP165	LPSSR023	PRG
2	Her4	A1	189191	293293	206206	116116	260264	139157
3	Her4	B1	185197	300300	208208	117117	281291	157159
4	Her4	C1	169189	293293	206206	117117	291252	159159
5	Her4	D1	169199	300300	208208	120120	291241	159159
6	Her4	E1	000000	291293	206206	117117	297256	153160
7	Her4	F1	191193	300300	208212	120126	260264	159159
8	Her4	G1	173199	293293	206206	120120	299317	157159
9	Her4	H1	189197	293293	206232	126131	297307	159159
10	Her4	A2	191197	293293	206232	120120	271297	159159
11	Her4	B2	179193	293293	206206	117120	317264	157159
12	Her4	C2	189189	293293	208208	120120	247247	139157
13	Her4	D2	169181	289293	206210	000000	291301	159159
14	Her4	E2	191197	300300	208212	117117	291301	139157
15	Her4	F2	179183	300300	208208	120126	297311	139157
16	Her4	G2	179169	000000	000000	000000	291260	139153
17	Her4	H2	189189	293293	206206	116129	297260	159159
18	Her4	A3	183197	291293	206206	117120	271313	139159
19	Her4	B3	179189	293293	206206	120120	331264	139159
20	Her4	C3	179193	293293	206226	117126	285260	139159
21	Her4	D3	181187	293293	206206	120120	000000	000000
22	Her4	E3	173199	300300	208212	120120	299299	139157
23	Her4	F3	189193	293293	206206	120129	260260	139139
24	Her4	G3	179185	293293	206206	120120	317323	000000
25	Her4	H3	189197	291293	206206	116116	299264	139139
26	Her4	A4	179197	293293	206236	116116	297297	159159
27	Her4	B4	179189	289293	206206	120120	273260	000000
28	Her4	C4	165239	291293	208210	117120	000000	139159
29	Her4	D4	000000	291293	206242	117117	297297	157157
30	Her4	E4	189239	300300	208208	000000	262269	139153
31	Her4	F4	185191	293293	206206	120120	297311	139159
32	Her4	G4	191163	000000	000000	000000	295311	139159
33	Her4	H4	185187	293293	206206	117117	277241	153159
34	Her4	A5	189199	293293	206210	120120	317247	159159
35	Her4	B5	189191	291293	206206	120120	297297	139159
36	Her4	C5	189189	293293	206206	131131	307252	153157
37	Her4	D5	183183	293293	206206	120120	000000	139159
38	Her4	E5	185191	293293	210232	122122	287317	157159
39	Her4	F5	179189	293293	208208	120120	297252	139159
40	Abe	a1	161189	293293	206208	120120	277315	157159
41	Abe	a2	189193	293293	206206	120120	273315	159159
42	Abe	a3	161161	293293	206206	120120	273315	157159
43	Abe	a4	161195	293293	206206	116116	301301	157157
44	Abe	a5	161161	293293	206208	000000	273273	000000
45	Abe	a6	161193	293293	206208	000000	273301	157159
46	Abe	b1	161193	293293	206208	000000	273273	159159
47	Abe	b2	161193	293293	206208	000000	273273	159159
48	Abe	b3	161161	293293	206206	120120	273252	157157

Prêt

démarrer pbarre sur 'Sa... SSR Eudora - [Entr... UGA



# Données de séquence de produit PCR

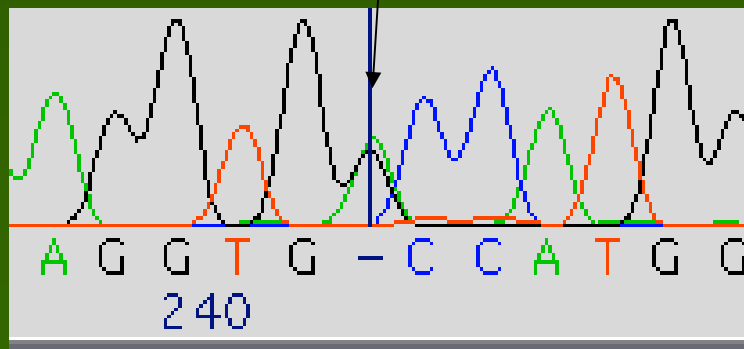


Microsoft Excel - SNPcommun\_10pc\_ss\_inde1\_ssSTS5\_STS5bis.xls

	A	B	C	D	E	F	G	H	I	J
1	pop	individu	STS4_60	STS4_113	STS4_226	STS4_386	STS4_475	STS5_173	STS5_490	STS5_594
68	Bre	BresD11	T	C	N	N	N	N	T	C
69	Bre	BresD12	N	C	T	T	T	N	N	Y
70	Bre	BresD7	N	Y	Y	Y	W	G	C	Y
71	Bre	BresD8	T	C	N	N	N	G	C	Y
72	Bre	BresD9	N	C	T	T	T	N	Y	Y
73	Bre	BresE10	T	C	T	T	T	N	N	T
74	Bre	BresE11	T	C	T	T	T	N	Y	T
75	Bre	BresE12	N	Y	Y	Y	W	N	T	Y
76	Bre	BresE7	T	C	T	T	T	N	T	Y
77	Bre	BresE8	T	C	T	T	T	G	C	T
78	Bre	BresE9	N	C	N	N	N	G	Y	T
79	Bre	BresF10	C	T	Y	C	A	A	T	C
80	Bre	BresF11	Y	Y	Y	Y	W	G	T	T
81	Bre	BresF12	T	C	T	T	T	N	N	Y
82	Bre	BresF7	T	C	T	T	T	N	N	Y
83	Bre	BresF8	T	C	T	T	T	N	T	Y
84	Bre	BresF9	Y	Y	Y	Y	W	N	N	Y
85	Bre	BresG10	T	C	T	T	T	N	N	Y
86	Bre	BresG11	N	Y	Y	Y	W	N	N	Y
87	Bre	BresG12	T	C	T	T	T	N	T	Y
88	Bre	BresG7	Y	Y	Y	Y	W	N	N	Y
89	Bre	BresG8	N	C	T	T	T	N	N	T
90	Bre	BresG9	N	C	T	T	T	N	C	Y
91	Bre	BresH10						N	T	Y
92	Bre	BresH11						N	T	Y
93	Bre	BresH7								
94	Bre	BresH8	N	Y	Y	Y	W	N	T	Y
95	Bre	BresH9	N	Y	Y	Y	W	N	T	Y
96	Cc	CcIA1	T	C	T	T	T	G	C	Y
97	Cc	CcIA2	N	C	T	T	T			
98	Cc	CcIA3	N	C	T	T	T			
99	Cc	CcIA4	N	C	T	T	T	A	T	C
100	Cc	CcIA5	N	C	T	T	T	N	C	Y

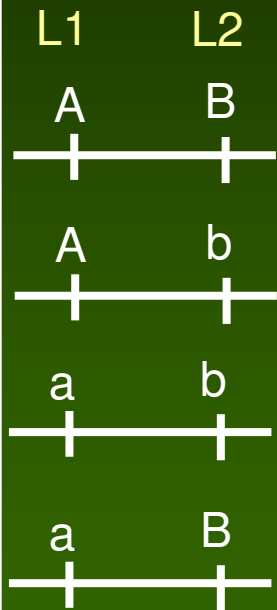
Microsoft Excel - SNPcommun\_10pc\_ss\_inde1\_ssSTS5\_STS5bis.xls

	B	C	D	E	F	G	H	I	J	K
E22										
1	individu	STS4_60	STS4_113	STS4_226	STS4_386	STS4_475	STS5_173	STS5_594	STS9_60	STS9_90
2	A1	002002	004004	002002	002002	002002	000000	000000	004004	004002
3	A2	000000	004004	002002	002002	002002	003003	002002	004002	004002
4	A3	002002	004004	002002	002002	002002	000000	004002	004002	002002
5	A4	000000	004004	002002	002002	002002	003003	002002	004002	002002
6	A5	002002	004004	002002	002002	002002	003003	002002	004002	002002
7	A6	002002	004004	002002	002002	002002	002002	004002	002002	002002
8	B1	000000	004002	004002	004002	001002	000000	004002	002002	002002
9	B2	000000	004004	002002	002002	002002	002002	000000	002002	002002
10	B3	000000	004004	002002	002002	002002	000000	002002	002002	002002
11	B4	002002	004004	002002	002002	002002	003003	002002	000000	000000
12	B5	004002	004002	004002	004002	001002	000000	002002	004002	002002
13	B6	000000	004004	002002	002002	002002	002002	000000	004002	002002
14	C1	000000	004004	002002	002002	002002	000000	002002	002002	002002
15	C2	000000	004004	002002	002002	002002	003003	002002	004002	002002
16	C3	002002	004004	002002	002002	002002	000000	000000	004002	004002
17	C4	002002	004004	002002	002002	002002	001002	000000	004002	004002
18	C5	002002	004004	002002	002002	002002	000000	004002	004002	002002
19	C6	004002	004002	004002	004002	001002	000000	002002	002002	002002
20	D1	000000	004002	004002	004002	001002	000000	000000	002002	000000
21	D2	000000	004004	002002	002002	002002	001003	004002	002002	002002
22	D3	000000	004004	002002	002002	002002	003003	002002	002002	002002
23	D4	002002	004004	002002	002002	002002	000000	004002	004002	002002
24	D5	002002	004004	002002	002002	002002	000000	002002	004002	002002



# Calcul du DL : Déséquilibre de liaison gamétique

Gamètes



Fréquences théoriques :

$H_0$  = association aléatoire  
entre allèles des 2 loci

		L1	
		A	a
L2	B	$f_A \times f_B$	...
	b	...	...

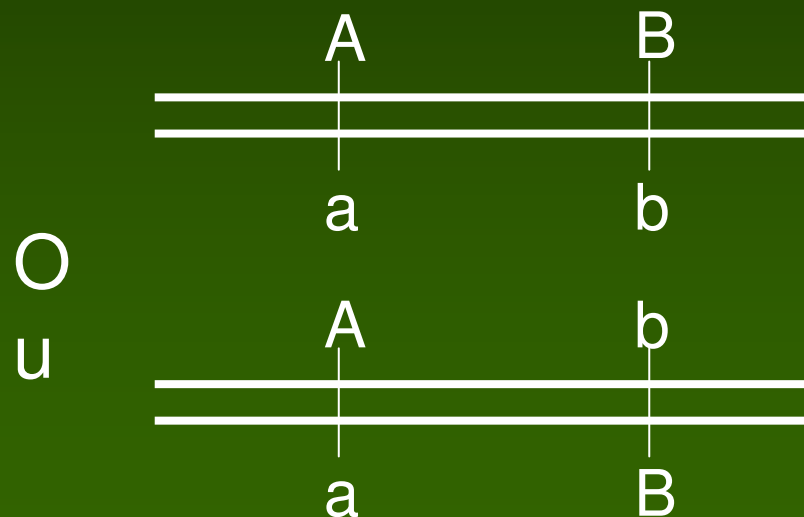
Fréquences observées :

		L1	
		A	a
L2	B	$f_{AB}$	...
	b	...	...

# Calcul du DL chez les hétérozygotes

- On ne connaît pas la phase haplotypique

Locus	L1	L2
Génotype	Aa	Bb



Solutions :

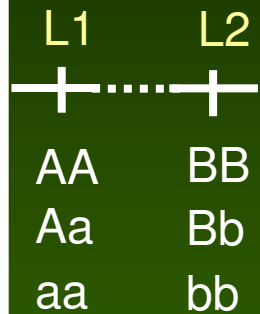
- Estimation de la phase en supposant l'équilibre d'Hardy

—Weinberg (logiciel PHASE 2.1  
Stephens et al. 2001, Stephens et  
Donnelly 2003)

- Calcul d'un déséquilibre de liaison génotypique

# Calcul du DL : Déséquilibre de liaison génotypique

Génotypes



Fréquences théoriques :

$H_0$  = association aléatoire  
entre génotypes des 2 loci

L2 \ L1	AA	Aa	aa
BB	...	...	...
Bb	...	$f_{Aa} \times f_{Bb}$	...
bb	...	...	...

Fréquences observées :

L2 \ L1	AA	Aa	aa
BB	...	...	...
Bb	...	$f_{Aa/Bb}$	...
bb	...	...	...

# Résultats

- Variabilité, équilibre HW et stucturation
- Déséquilibre de liaison



# Variabilité et équilibre HW

SSR	Abe	Her3	Cc
Nombre moyen d'allèles	3,3	10,3	16,3
Hétérozygotie attendue	0,42	0,60	0,76
Fis	0,14	0,06	0,26*

\* Significatif à 0,01

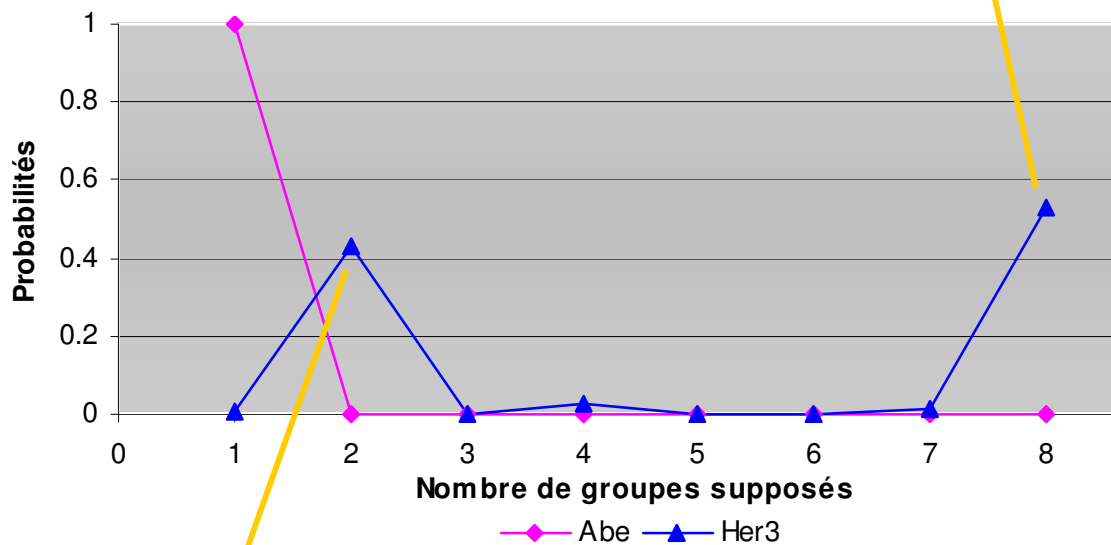
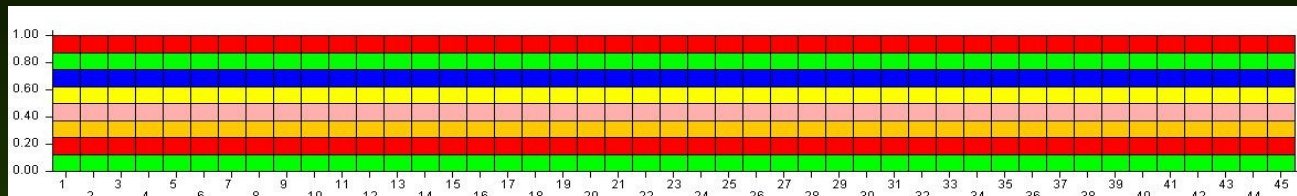
STS	Aber	Her3	Cc
pb / SNP	72	32	30
Fis	-0,03	-0,01	0,29*

$\pi$	:			
	STS4	0,0002	0,005	0,004
	STS5	0,0002	0,008	0,008

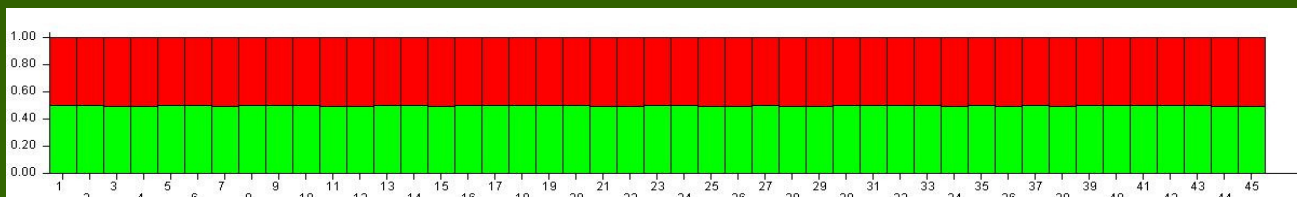
Conclusion :

- Aberavon très peu variable
- Variabilité d'Herbie comparable à celle de la Cc
- Pas d'écart à l'équilibre HW dans ces variétés synthétiques

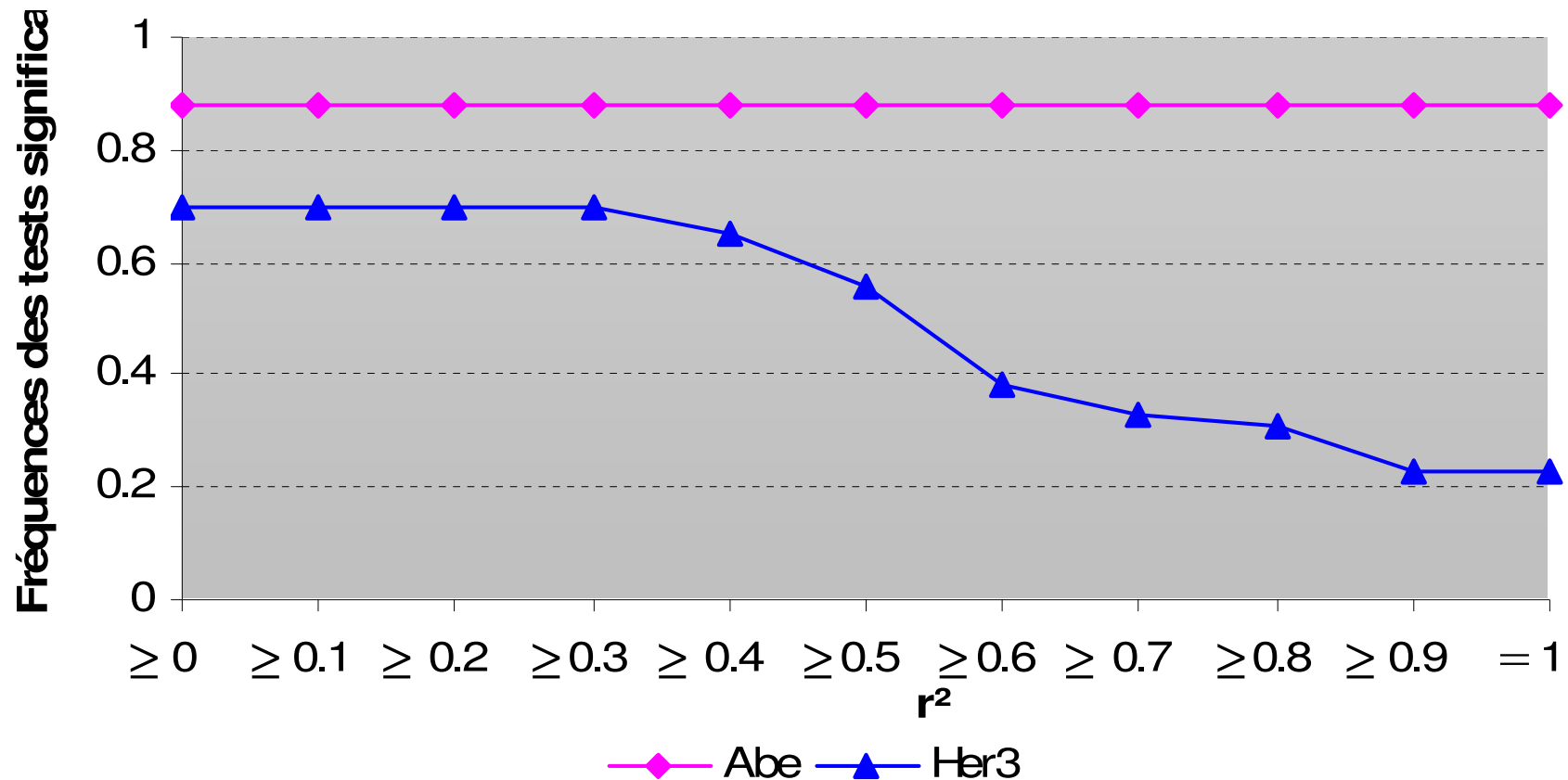
# Structuration



Conclusion :  
Pas de structuration  
dans ces variétés  
synthétiques

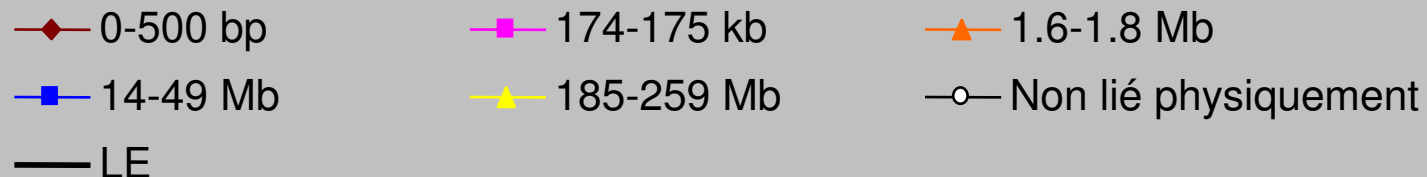
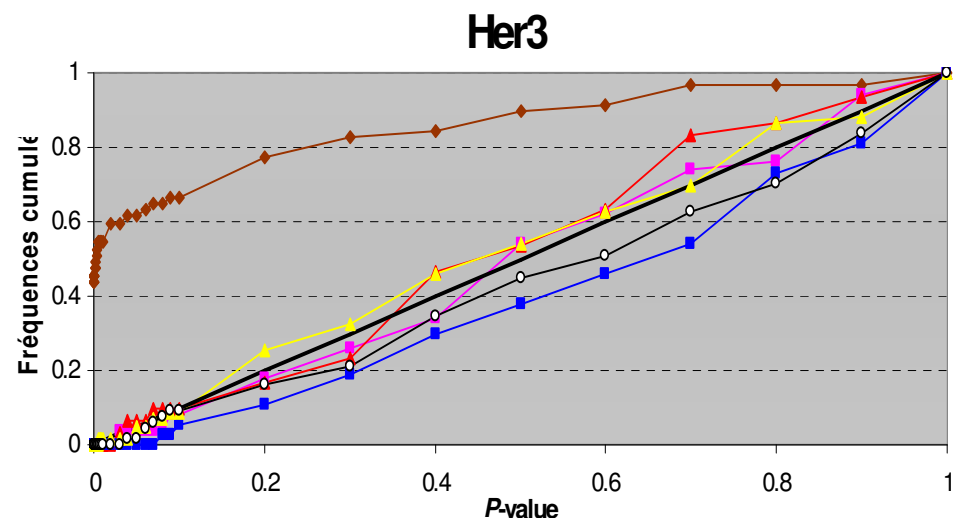
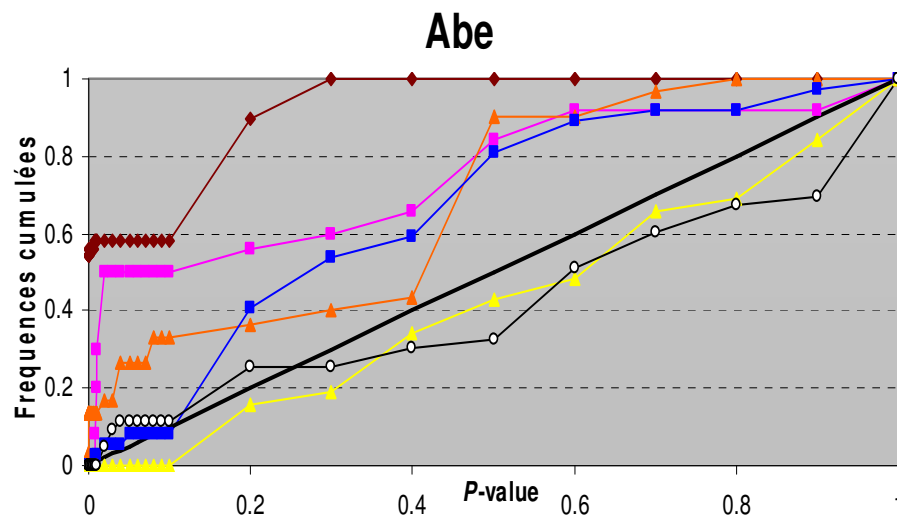


# Déséquilibre de liaison autour de GAI : DL gamétique à courte distance (0 – 800 pb)



Conclusion : Valeurs de DL Abe > Valeurs de DL Her3

# Déséquilibre de liaison autour de GAI : DL génotypique à grande distance



Conclusion : Le DL est significatif sur une plus grande distance dans Abe que dans Her3

# Conclusion Partie 1

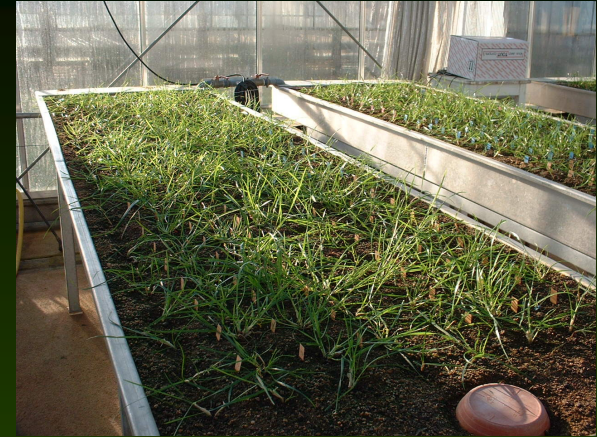
- La variabilité des variétés synthétiques est plus ou moins grande en fonction du nombre de constituants
- Pas de structuration et pas d'écart à l'équilibre HW
- Dans les variétés à base génétique large le DL décroît rapidement  $\Rightarrow$  "gène candidat"
- Dans les variétés peu variables le DL décroît lentement mais plus rapidement qu'attendu
  - ceci peut s'expliquer par l'appareillement des 6 constituants

## Partie 2 : Etude d'association entre GAI et la croissance foliaire

- Etude de type “gène candidat” dans une variété à base génétique large : Herbie
  - Variabilité phénotypique (croissance foliaire)
  - Variabilité génétique (gène GAI)
  - Recherche de corrélations entre polymorphisme phénotypique et polymorphisme génétique

# Phénotypage

2 expérimentations



Printemps

- 186 individus de Her4

Automne

- 186 individus de Her4
- 100 individus Cc

Cinétique de croissance  
sur les feuilles entières  
3 (f3) et 4 (f4)



Longueur maximale :  $L_{max}$

Vitesse maximale :  $V_{max}$   
d'élongation

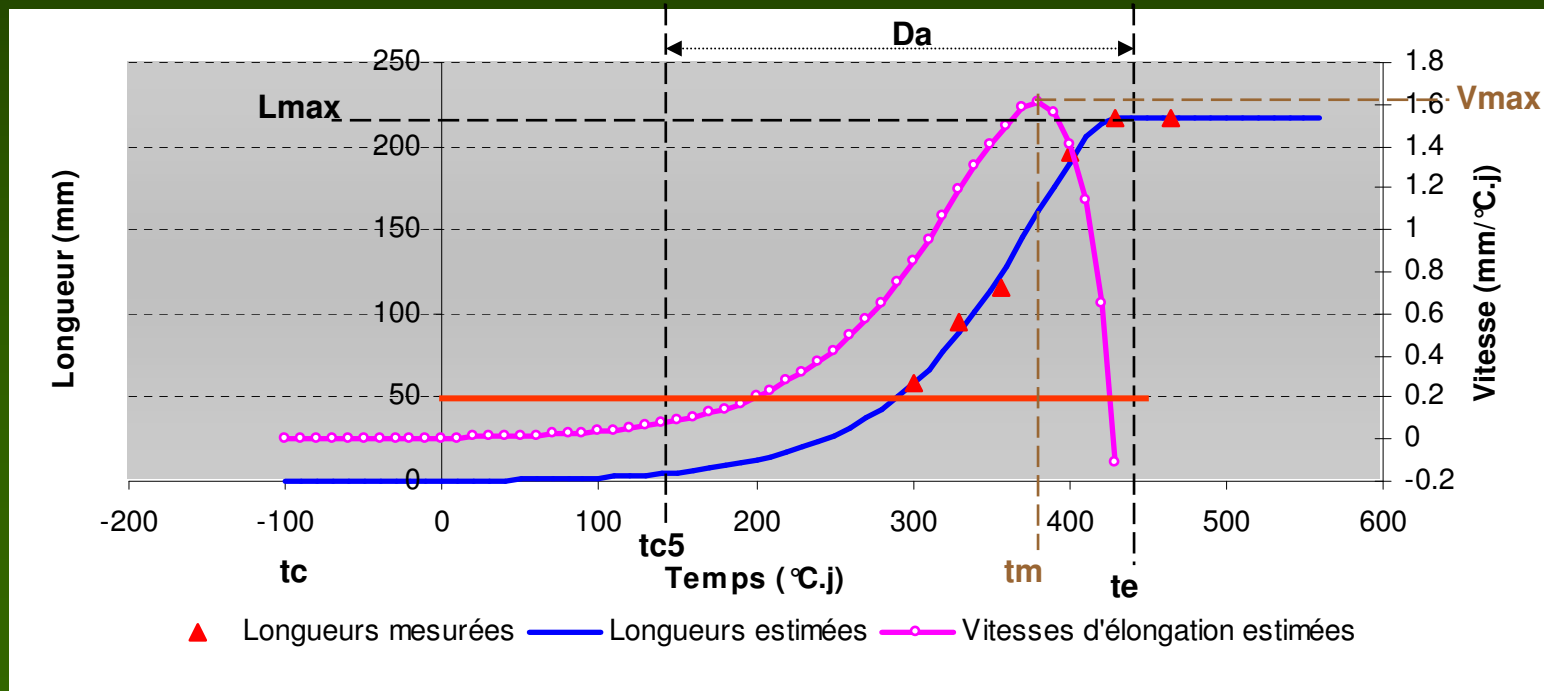
Durée d'allongement :  $D_a$

# Modélisation de la croissance foliaire :

## Fonction Béta

Longueur fonction de  $t$ ,  $t_c$ ,  $t$  (vitesse maximale),  $t_e$  et longueur maximale

(Verdenal et al. in press)





# Variabilité de la croissance foliaire

- Données printemps et automne analysées séparément car :
  - Il existe des interactions génotype X période
  - Il existe des QTL de longueur de feuilles différents au printemps et à l'automne

# Résultats

- Variabilité de la croissance foliaire
- Variabilité du gène GAI
- Etude d'association GAI / croissance foliaire

## Variabilité de la croissance foliaire : feuille 3, Her4

	Printemps		Automne	
Variable	Effet géno. (p)	Héritabilité individuelle	Effet géno. (p)	Héritabilité individuelle
Lmax	0,0001	0,77	0,0001	0,64
Vmax	0,0001	0,74	0,0001	0,67
Da	0,0001	0,46	0,0001	0,39

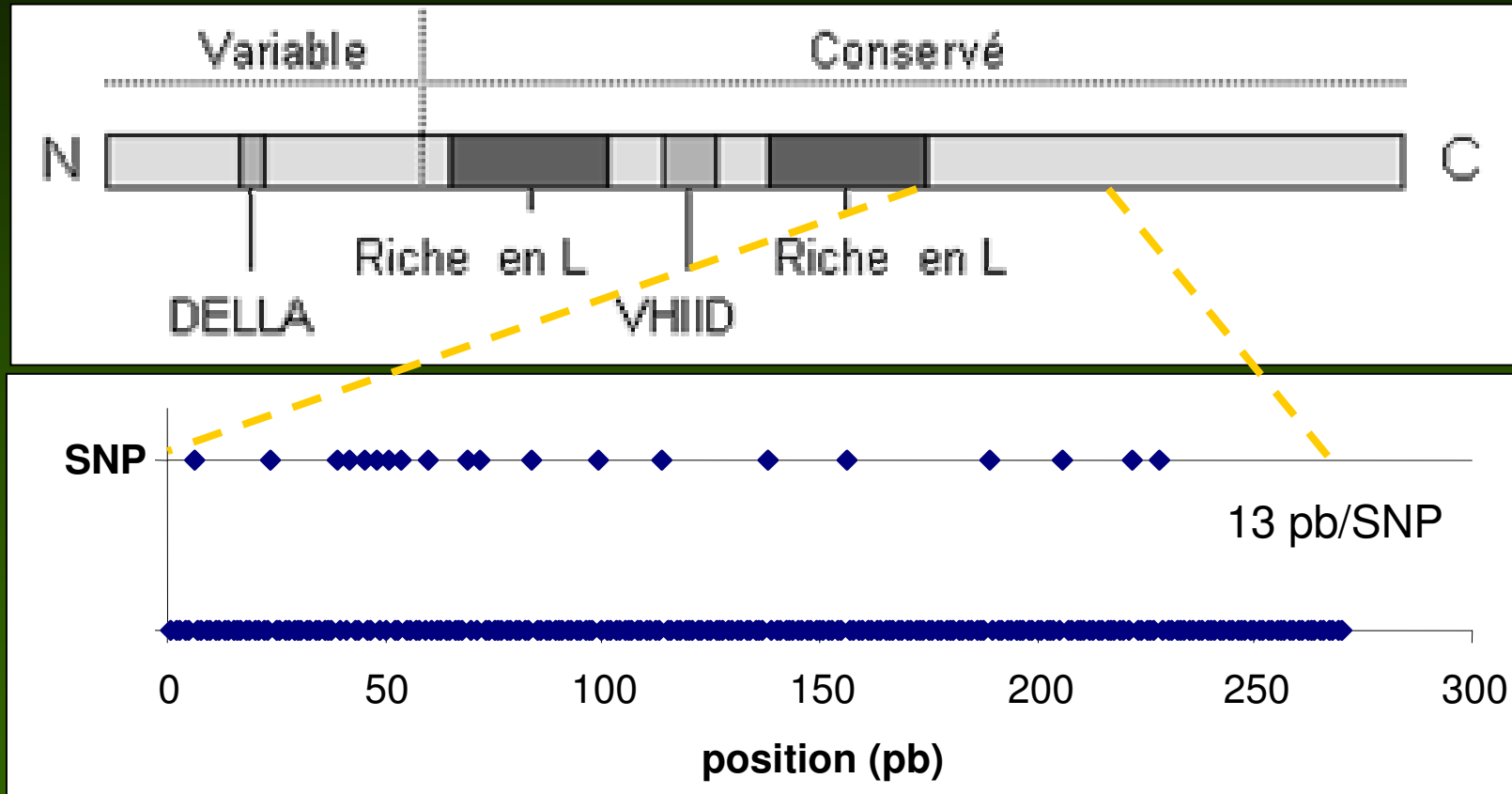
Conclusion : fortes héritabilités pour la longueur et la vitesse maximale et plus faible héritabilité pour la durée d'allongement

## Variabilité de la croissance foliaire : feuille 3, automne, Her4 et Cc

Variable	Max	Min	Moy	Var
Her4 Lmax (mm)	450	126	301	4245
Cc Lmax (mm)	478	118	270	4814
Her4 Vmax (mm/°Cj)	2,56	0,84	1.64	0,09
Cc Vmax (mm/°Cj)	2,73	0,81	1,57	0,13
Her4 Da (°Cj)	401	187	267	909
Cc Da (°Cj)	321	173	245	841

Conclusion : La variabilité phénotypique d'Her4 est comparable à celle de la Cc

# Variabilité du gène GAI chez Her4



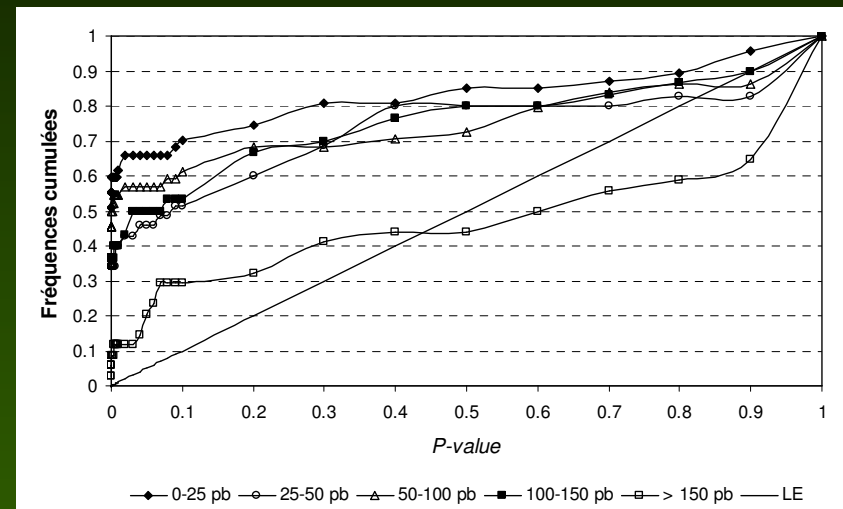
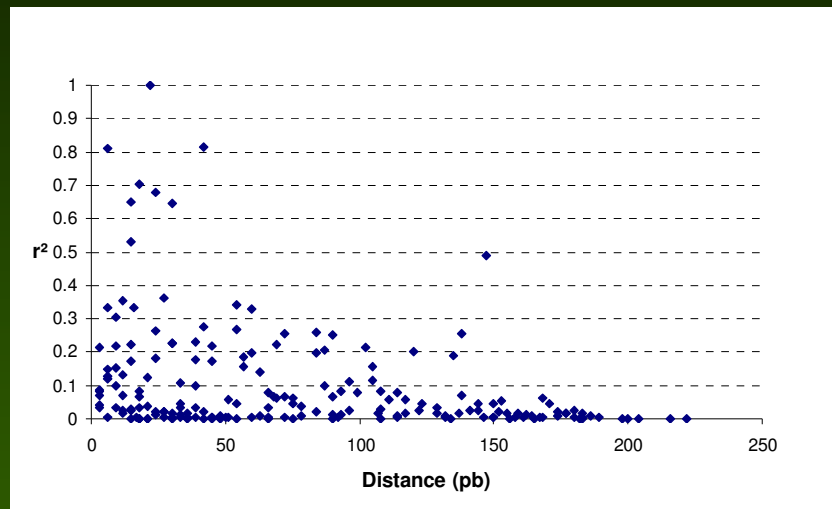
# Variabilité du gène GAI chez Her4

39 haplotypes reconstitués à l'aide de PHASE dont 9 majoritaires

Haplotype no.	Haplotype	Inferred no. in sample
2	CCACTGAAG <u>C</u> GGCGCGCAG	66
9	CCGACGGGG <u>C</u> GGCGCCCAG	57
10	CCGACGGGG <u>C</u> CGCGCGCAG	17
15	CCGACCAGG <u>C</u> GGCGCGCAG	14
24	CCGCCGAAGG <u>C</u> CCCACGGCAG	56
25	CCGCCGAAGG <u>C</u> CCCACGGCGG	15
27	CCGCCGAGCG <u>C</u> CCC CGCGCAG	50
30	CCGCCGGAGGGG <u>C</u> CGCGCAG	18
37	CCGCCCGGGG <u>C</u> CCC CGCGCAG	26



# Déséquilibre de liaison dans GAI chez Her4



Conclusion : Dans GAI DL significatif jusqu'à 150 pb

# Etude d'association

- Trois méthodes :
  - Analyse de variance sur la présence versus absence de chaque haplotype
  - Régression linéaire multiple en considérant tous les SNP
  - Analyse de variance sur les groupes formés à partir d'un arbre phylogénétique (TREESCAN, Templeton et al., 2005)



# Analyse de variance sur la présence versus l'absence de chaque haplotype

Haplo.	Nb. of samples +	Nb. of samples -	Printemps			Automne		
			Lmax	Vmax	Da	Lmax	Vmax	Da
2	60	126	<b>0,032</b>	0,071	0,169	0,649	0,249	0,208
9	50	136	0,810	0,645	0,220	0,684	0,143	0,130
10	17	169	0,491	0,467	0,706	0,451	0,681	0,355
15	14	172	0,953	0,603	0,761	0,331	0,185	0,600
24	54	132	0,786	0,496	0,970	0,594	0,322	0,283
25	15	171	0,623	0,998	0,725	0,511	0,830	0,369
27	45	141	0,077	0,123	0,410	0,227	0,286	0,801
30	17	169	0,831	0,418	0,066	0,848	0,918	0,756
37	25	161	0,912	0,978	0,758	0,671	0,525	0,943

# Régression linéaire multiple sur les 20 SNP

## Printemps

Variables	SNP	R <sup>2</sup> partiel	R <sup>2</sup> global
<b>Lmax</b>	<b>GAI069</b>	<b>0,0502</b>	0,0934
(mm)	GAI060	0,0255	
	GAI039	0,0177	
<b>Vmax</b>	<b>GAI069</b>	<b>0,0407</b>	0,1008
(mm/°Cj)	GAI099	0,0233	
	GAI206	0,0185	
	GAI189	0,0183	
<b>Da</b>	<b>GAI069</b>	<b>0,0187</b>	0,0755
(°Cj)	GAI060	0,0164	
	GAI039	0,0159	
	GAI048	0,0124	

## Automne

Variables	SNP	R <sup>2</sup> partiel	R <sup>2</sup> global
<b>Lmax</b>	<b>GAI069</b>	<b>0,038</b>	0,1106
(mm)	GAI099	0,0295	
	GAI048	0,0252	
	GAI222	0,018	
<b>Vmax</b>	<b>GAI069</b>	<b>0,0418</b>	0,1358
(mm/°Cj)	GAI048	0,0219	
	GAI114	0,018	
	GAI156	0,0174	
	GAI099	0,0122	
	GAI189	0,0245	
<b>Da</b>	GAI051	0,0341	0,0838
(°Cj)	<b>GAI069</b>	<b>0,0225</b>	
	GAI114	0,0144	
	GAI048	0,0128	

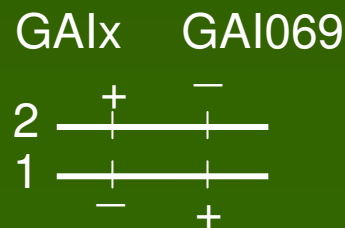
# Etude d'association :

## Effet du SNP GAI069

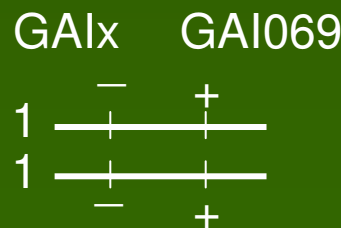
Variables	Moyenne	Effet des génotypes par rapport à la moyenne		
		1 / 1	2 / 2	1 / 2
Lmax_pts	312	5	-27	15
Lmax_aut	303	-6	-17	13
Vmax_pts	1,732	0,038	-0,122	0,059
Vmax_aut	1,507	0,081	-0,019	0,047
<b>Da_pts</b>	<b>273</b>	<b>1</b>	<b>-6</b>	<b>9</b>
<b>Da_aut</b>	<b>297</b>	<b>-5</b>	<b>-6</b>	<b>5</b>

Conclusion :

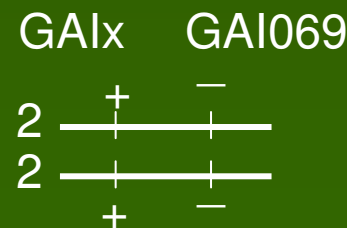
Les hétérozygotes sont toujours supérieurs aux homozygotes



>



>

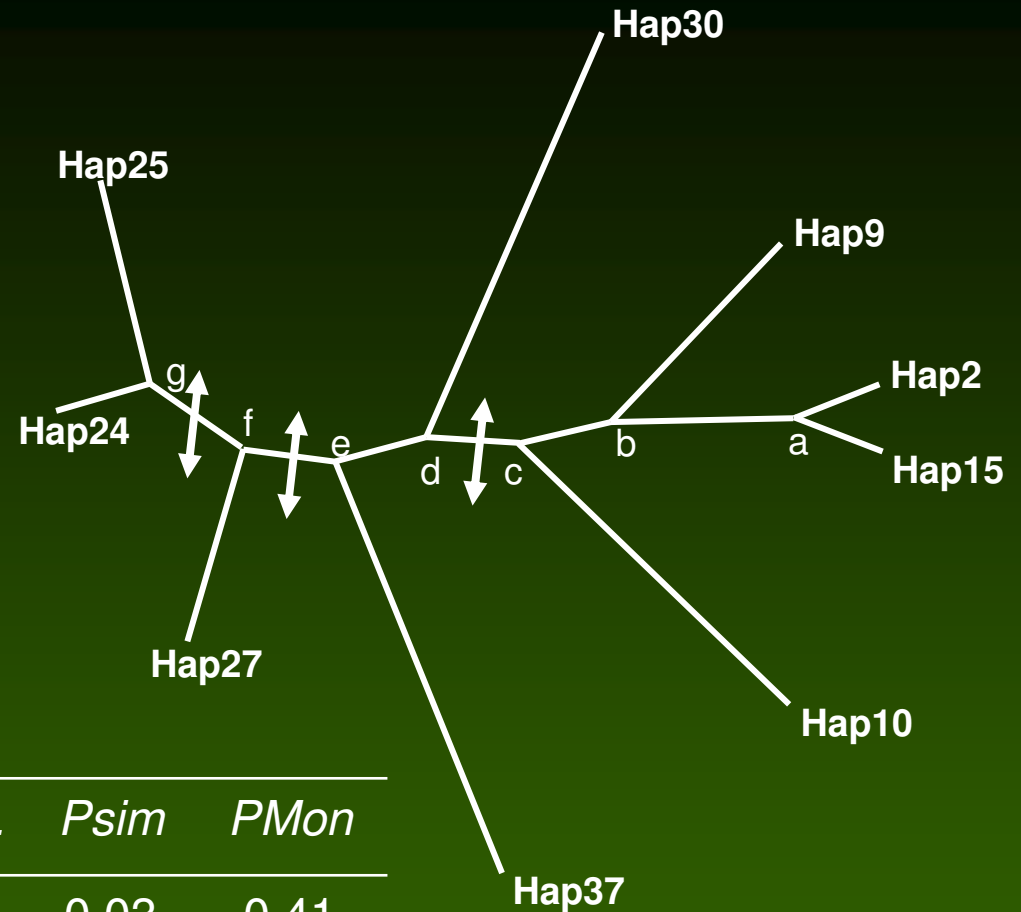


Avec :

GAIx < GAI069

+ dominant

# Analyse de variance sur les groupes basés sur un arbre phylogénétique



Trait	Branch	<i>F-stat</i>	% var.	<i>Psim</i>	<i>PMon</i>
Lmax pr	c-d	3.7	0.04	0.03	0.41
Vmax pr	c-d	5.1	0.06	<b>0.007</b>	0.16
Vmax pr	f-g	4.1	0.04	0.02	0.30
Vmax pr	e-f	3.2	0.03	0.04	0.50
Vmax aut	f-g	3.3	0.04	0.04	0.52



# Etude d'association : comparaison des 3 méthodes

- Toutes ont tendance à faire ressortir l'effet de GAI069  
Mais la régression sur les SNP semble plus puissante
  - Pas de bruit de fond dû à des SNP ayant des effets contraires présents dans le même haplotype
- L'analyse de variance sur les haplotypes peut être intéressante si l'effet du gène est provoqué par plusieurs mutations
- L'analyse basée sur l'arbre phylogénétique est intéressante si, au cours de l'évolution, il y a eu des mutations reverses ou des branches de l'arbre dans lesquelles il n'y a pas de DL entre le SNP étudié et celui à l'origine de l'effet

## Conclusion Partie 2

- GAI explique environ 10% des variations de longueur de feuilles dans la variété Herbie malgré le grand nombre de gènes candidats
- Il intervient à la fois sur la durée et la vitesse d'allongement foliaire
- Un SNP explique jusqu'à 5% de la variabilité des variables de longueur de feuilles étudiées
- Les SNP étudiés sont synonymes donc non directement impliqués

# Conclusion générale

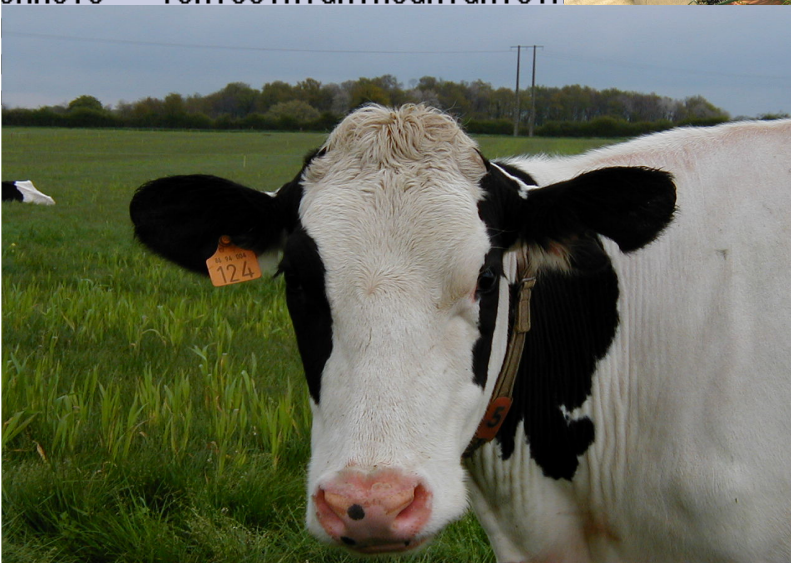
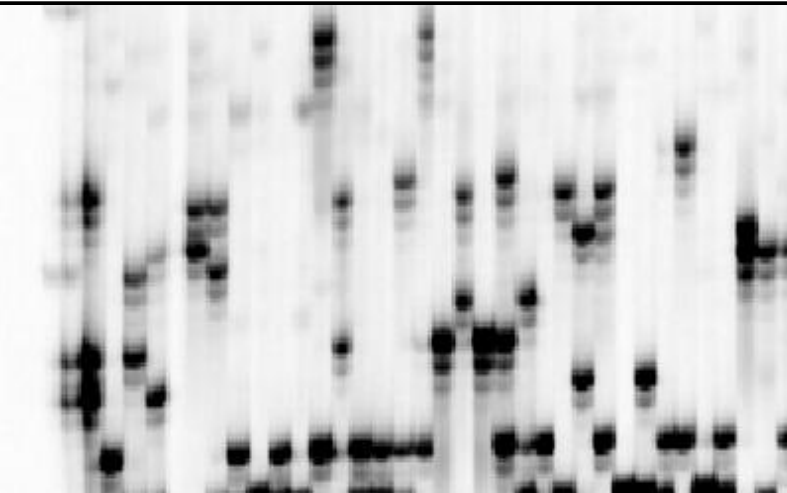
- Les variétés synthétiques peuvent être utilisées pour les études d'association
  - “gène candidat” dans les variétés à base large
  - “genome scan” dans les variétés à base étroite mais dans une région ciblée
- Nécessité de mesurer le DL dans la population et dans la région étudiée

# Perspectives

- Obtenir la séquence complète du gène GAI pour identifier la ou les mutations responsables des variations observées
- Etudier l'effet de ce gène dans d'autres populations
- Possibilité d'intégrer les études d'association dans les schémas de sélection en particulier pour les synthétiques



CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA  
CAACTC\*\*\*TCATCCTATGATACGATGATCTA



**Merci à tous ...**

# Références bibliographiques

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357-374

Gupta PK, Sachin R, Kulwal P (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461-485

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* 67:170-181

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162-1169

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978-989

Verdenal A., Combes D., Escobar-Gutiérrez A.J. 2008. A study of ryegrass architecture as a self-regulated system, using functional-structural plant modeling. *Functional Plant Biology*, 35 (9) (in press).

# Logiciels

- STADEN : vérification et alignement des séquences (<http://staden.sourceforge.net/> )
- BLAST sous NCBI : vérification de l'identité de la séquence par homologie avec d'autres connues (<http://blast.ncbi.nlm.nih.gov/Blast.cgi> )
- DNAsp et TASSEL : calcul et test du DL gamétique (<http://www.ub.edu/dnasp/> et <http://www2.maizegenetics.net/index.php?page=bioinformatics/tassel/index.html> ) en plus calcul de l'indice de diversité  $\pi$  avec DNAsp
- PHASE : détermination de la phase entre locus à partir de données génotypiques (<http://www.stat.washington.edu/stephens/phase.html>)
- GENETIX : calcul du nombre d'allèle, de l'hétérozygotie attendue et de l'indice de fixation Fis (<http://www.genetix.univ-montp2.fr/genetix/intro.htm>)
- GENEPOP : test l'écart à l'équilibre d'Hardy-Weinberg, test du DL génotypique (<http://genepop.curtin.edu.au/>)
- STRUCTURE : évaluation de la structuration des populations (<http://pritch.bsd.uchicago.edu/structure.html> )
- SAS : analyses de variance, corrélations, régressions linéaire (<http://www.sas.com/technologies/analytics/statistics/stat/index.html> )
- PHYLIP : détermination d'un arbre phylogénétique à partir des données de séquences (<http://evolution.genetics.washington.edu/phylip.html> )
- Phylodraw : représentation graphique d'un arbre phylogénétique (<http://pearl.cs.pusan.ac.kr/phylodraw/> )
- TREESCAN : recherche d'associations en prenant en compte l'arbre phylogénétique (<http://darwin.uvigo.es/software/treescan.html> )