

Développement d'un outil pour la sélection assistée par marqueurs chez le ray-grass anglais

Philippe Barre

INRA Centre Poitou-Charentes

UR4 : Unité de Recherche Pluridisciplinaire, Prairies et Plantes Fourragères
(URP³F)



Séminaire CTPS 14 mars 2013

Contexte

- Les prairies représentent **45% de la SAU** française (14 Millions d'ha) → rôle majeur dans l'alimentation des herbivores et dans d'autres services éco-systémiques (qualité de l'eau, stockage de C...)
- Une partie des prairies utilisée dans les systèmes prairiaux les plus intensifs est semée régulièrement: **prairies temporaires** recouvrant 2,7 Millions d'ha
- Les prairies temporaires sont semées avec du **matériel sélectionné** pour différents caractères dont le rendement, la résistance aux maladies, la qualité...



Contexte

- Le **ray-grass anglais** (*Lolium perenne* L.) est l'espèce de graminées fourragères la plus semée dans les régions tempérées
- C'est une espèce :
 - diploïde **2n=14** avec un génome de taille moyenne 1C=2.3 Gb
 - très polymorphe : 1 SNP (Single Nucleotide Polymorphism) / 15-30 bases
 - présentant un système d'auto-incompatibilité et une forte dépression de consanguinité

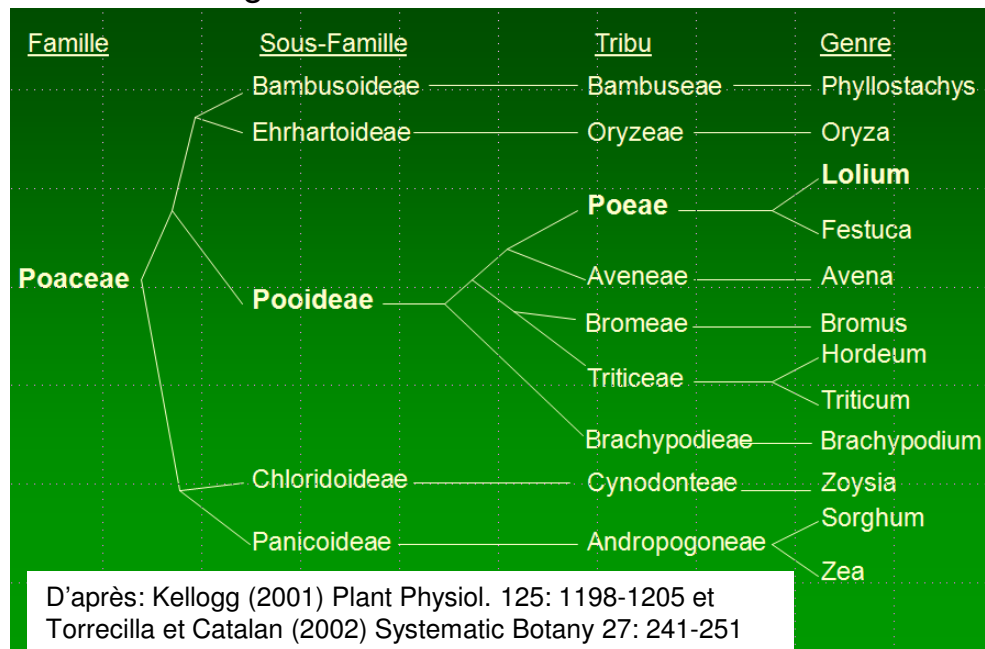
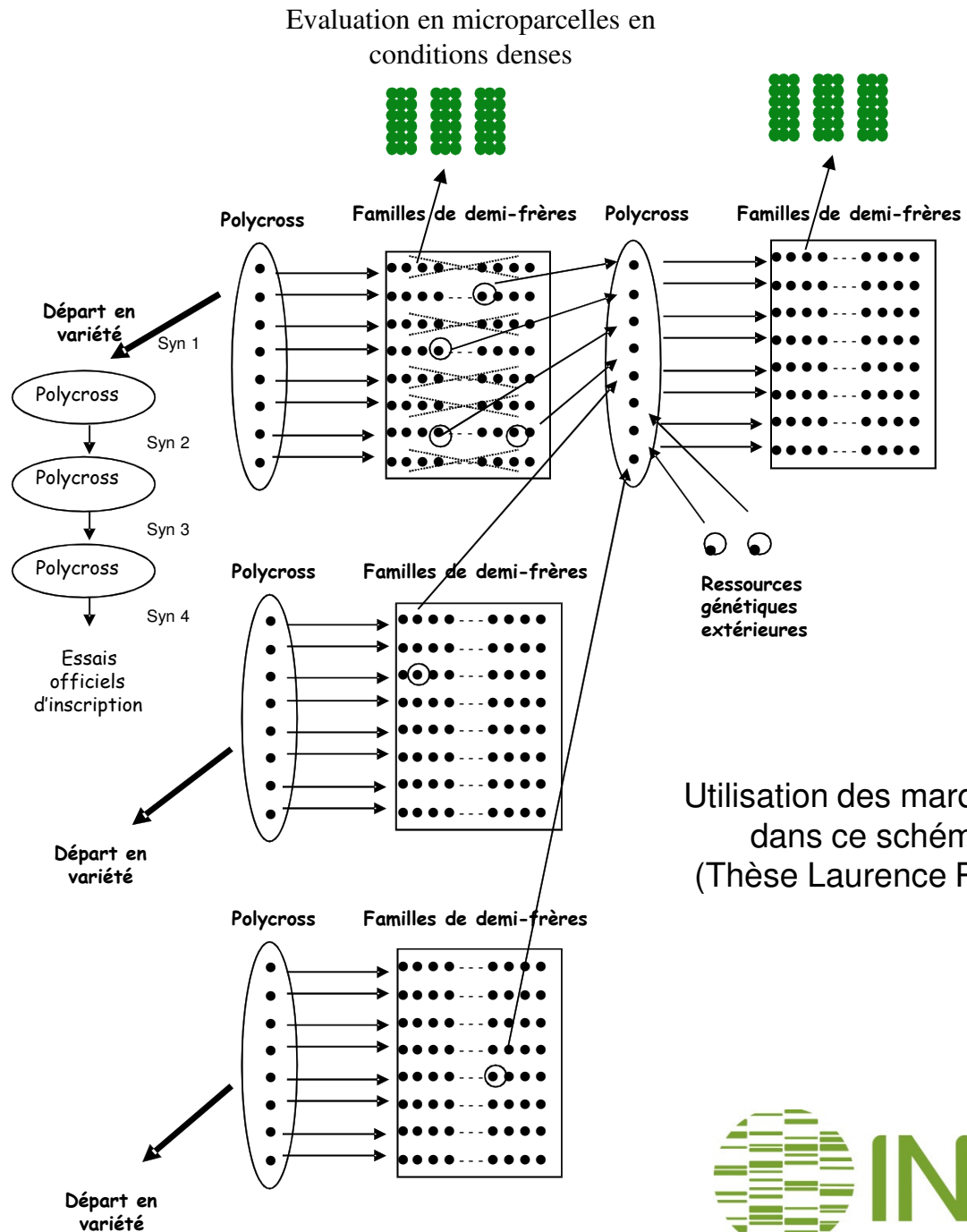


Schéma de création variétale du ray-grass anglais

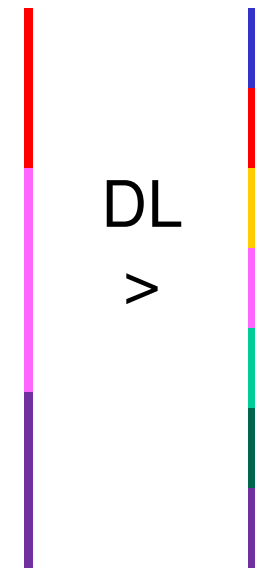
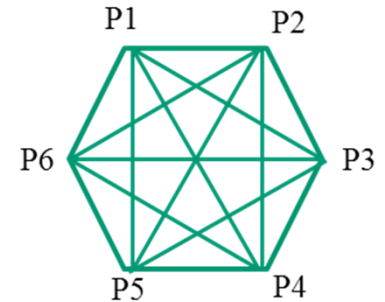


Variétés
synthétiques



Contexte

- Les descendance de polycross sont des **familles de pleins-frères connectées** qu'il est possible d'identifier grâce aux marqueurs moléculaires : tests de paternité
- Ces familles ont un **déséquilibre de liaison important**
- Dans ces familles, une centaine de marqueurs bien répartis sont suffisants pour couvrir l'ensemble du génome
- Ces marqueurs doivent être **spécifiques de locus** pour pouvoir comparer les cartes entre les parents afin d'identifier les allèles les plus favorables (logiciel MCQTL)
- **Faible coût** du génotypage pour un grand nombre d'individus « **haut débit** » : marqueurs SNP



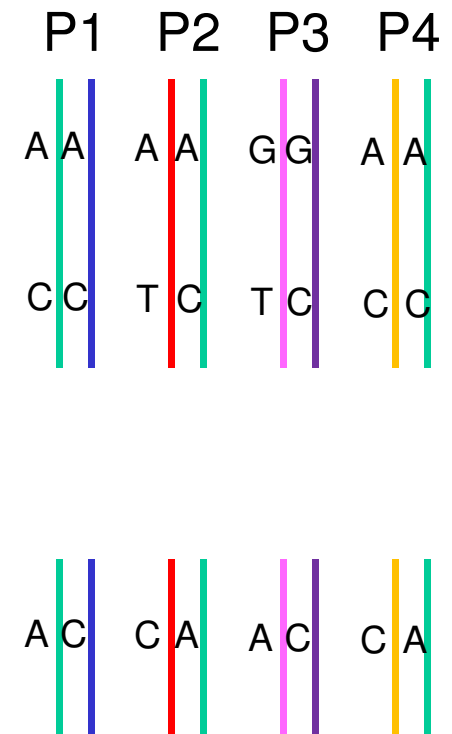
Objectif

Développer un outil permettant
d'obtenir une centaine de marqueurs
SNP bien répartis sur le génome chez
le ray-grass anglais

Stratégie

- **Amplifier des fragments** d'ADN bien répartis sur le génome pour chaque parent des polycross d'intérêt
- **Séquencer les amplicons** en utilisant les nouvelles technologies de séquençage en marquant chaque parent (Tag)
- **Détecter les SNP** intra- et inter- parents
- **Développer** des marqueurs SNP puis de **génotyper** les descendance de polycross

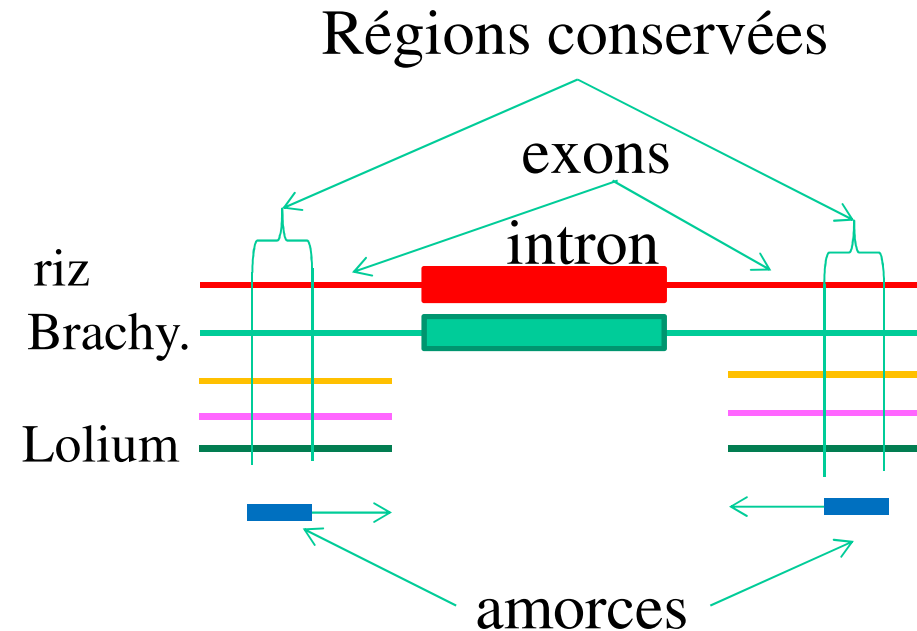
Chromosome 1



Définition d'amorces

But :

- Amorces dans des **régions conservées** de chaque coté d'un intron
- Taille d'amplification de 300-600 bp
- Bien réparties sur le génome



- Séquences d'EST de *Lolium* et *Festuca* disponibles sur la base de données NCBI
- EST de *Lolium* cartographiés (Miura et al. 2007)
- Séquence génomique de *Brachypodium distachyon*

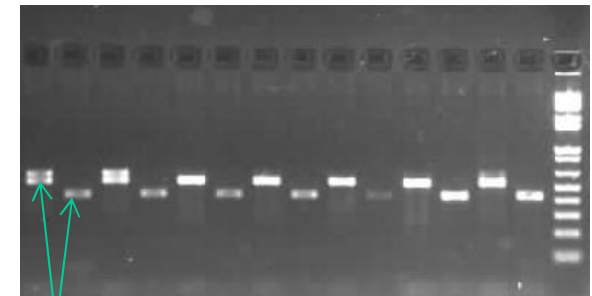
Définition d'amorces

- Manuellement avec Oligo 6 : **90 couples d'amorces**
 - Alignement d'une séquence de *Lolium* avec au moins une autre séquence de graminée
- Logiciel COS-finder (Quraishi et al. 2009): **432 couples d'amorces**
 - Définition des amorces sur les EST de *Lolium*
- En plus, **144 couples d'amorces** tirés de Tamura et al. 2009 mais pas d'information de position

Total : 666 couples d'amorces

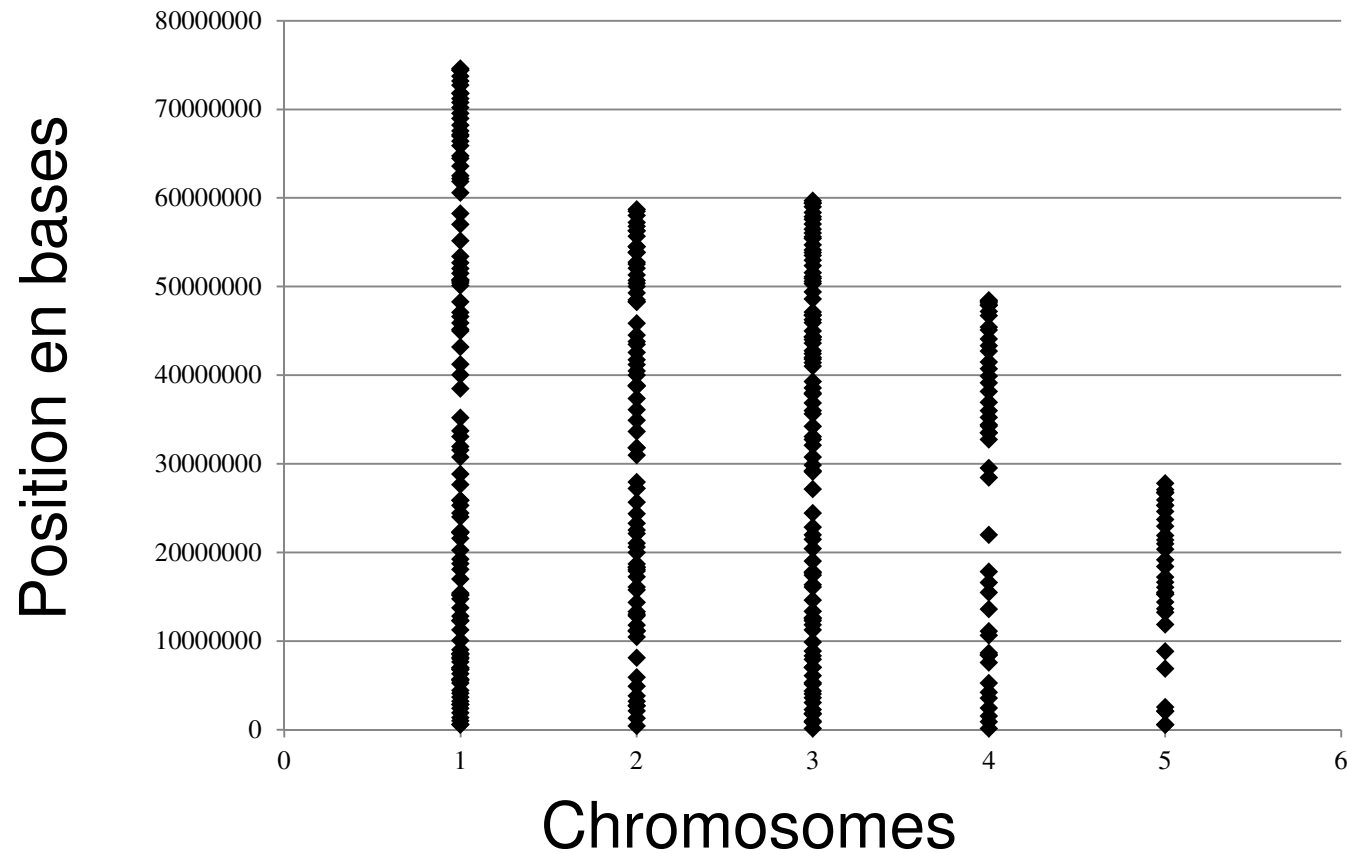
Test d'amplification chez le ray-grass anglais

- Sur 7 génotypes :
 - Aberavon, Carillon, Vigor, Gagny: **fouillage**
 - Marathon, Transate, Concerto: **gazon**
 - Migration sur gel d'agarose
 - Sélection des couples d'amorces qui révèlent :
 - Au moins une bande sur chaque génotype
 - Pas plus de deux bandes par génotype
- 363 couples d'amorces



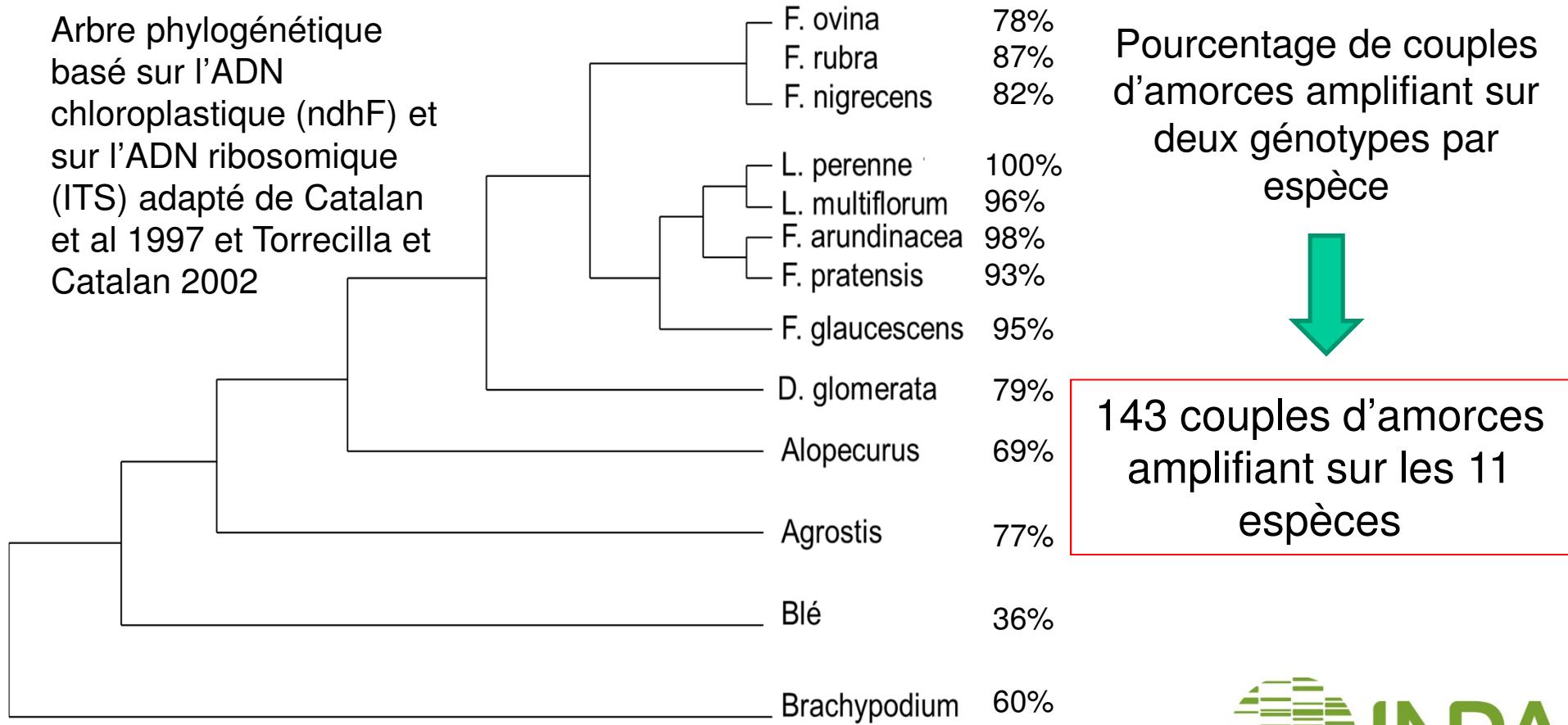
2 couples d'amorces sur 7 génotypes

Distribution des 363 couples d'amorces sur le génome de *Brachypodium*



Portabilité à d'autres espèces prairiales: multispecies

- Test d'amplification sur 11 espèces de graminées prairiales



Séquençage de 346 amplicons de 7 géotypes de ray-grass anglais

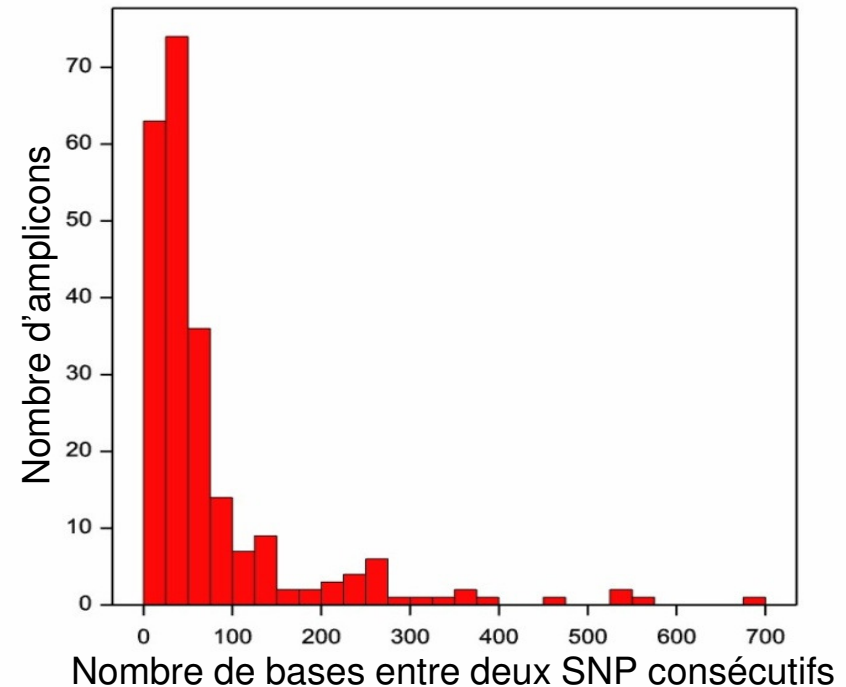
But: profondeur de séquençage de 60 lectures (double sens)/amplicon/géotype 2x
soit 41520 lectures / géotype

Varieties	Méthode de séquençage	Nombre de lectures	Nombre d'amplicons séquencés
Gagny	1/8 of 454 GS FLX (87500 lect.)	61784	203
Transate	1/8 of 454 GS FLX	58379	203
Aberavon	2/3 of 454 GS junior (46660 lect.)	21643	125
Concerto	2/3 of 454 GS junior	22780	157
Vigor	2/3 of 454 GS junior	23475	150
Marathon	2/3 of 454 GS junior	14956	105
Carillon	2/3 of 454 GS junior	17386	83

- Manque de profondeur
- Séquençage différentiel des amplicons

Polymorphisme sur 251 amplicons (94114 bp) chez le ray-grass anglais

- Polymorphisme intra- et inter-génotypes : 2819 SNP
- Polymorphisme chez tous les amplicons sauf un
- Le nombre de base moyen entre deux SNP consécutifs de 42



Fort polymorphisme même
sur un nombre limité de
génotypes

Niveau d'homozygotie

Variétés	Nombre d'amplicons homozygotes	Pourcentage d'amplicons homozygotes
Aberavon	69	55
Carillon	50	60
Concerto	90	57
Gagny	101	50
Marathon	62	59
Transate	143	70
Vigor	85	57
Total	600	58

- Le nombre d'amplicons homozygotes s'est révélé légèrement supérieur à l'attendu chez le ray-grass anglais ($H_o = 0.4-0.6$ soit de 60 à 40% d'homozygotes)
- Hypothèses:
 - Quelques allèles avec des fréquences élevées dans chaque variété
 - Homogamie et consanguinité plus fortes qu'attendu

SNP hétérozygotes

Pourcentage de SNP hétérozygotes chez différents génotypes

Nombre de génotypes	Pourcentage de SNP
0	19
1	42
2	21
3	11
4	2
5	4
6	0
7	0

19% des SNP ont révélé du polymorphisme qu'entre individus

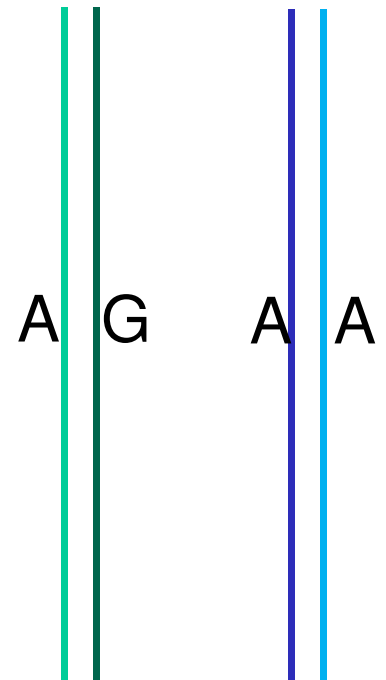
P1	P2	P3	P4	P5	P6	P7
AA	AA	GG	AA	AA	AA	GG
AA	AA	GA	AA	AA	AA	GG
CC	TC	TT	CC	TC	TC	CC
GA	AA	GG	GA	AA	GA	GA

- Pour un locus donné, plusieurs SNP seront nécessaires pour génotyper tous les individus
- Il est nécessaire de séquencer les génotypes d'intérêt pour identifier les SNP informatifs

Génotypage

- Identification d'une centaine de SNP « génotypables » sur une descendance (Aberavon x Gagny)
- Peu de SNP « idéaux » : amplicons de 500 – 600 bases plus informatifs
- Sur 32 marqueurs SNP, 28 ont donné de bon résultats avec la méthode Kaspar

Aberavon Gagny



50 bases

- Pas d'insertion / délétion
- Le minimum de polymorphisme
- Eviter les polyN

Conclusions

- **363 couples d'amorces** répartis sur le génome sont disponibles chez le ray-grass anglais et largement transférables aux fétuques et au dactyle
- Le logiciel **COS-finder** s'est révélé très efficace pour la définition d'amorces
 - Possibilité d'augmenter le nombre de marqueurs dans une région spécifique en se basant sur la position chez *Brachypodium* et bientôt chez le ray-grass directement
 - Cibler une taille d'amplicons proche de 500 bp

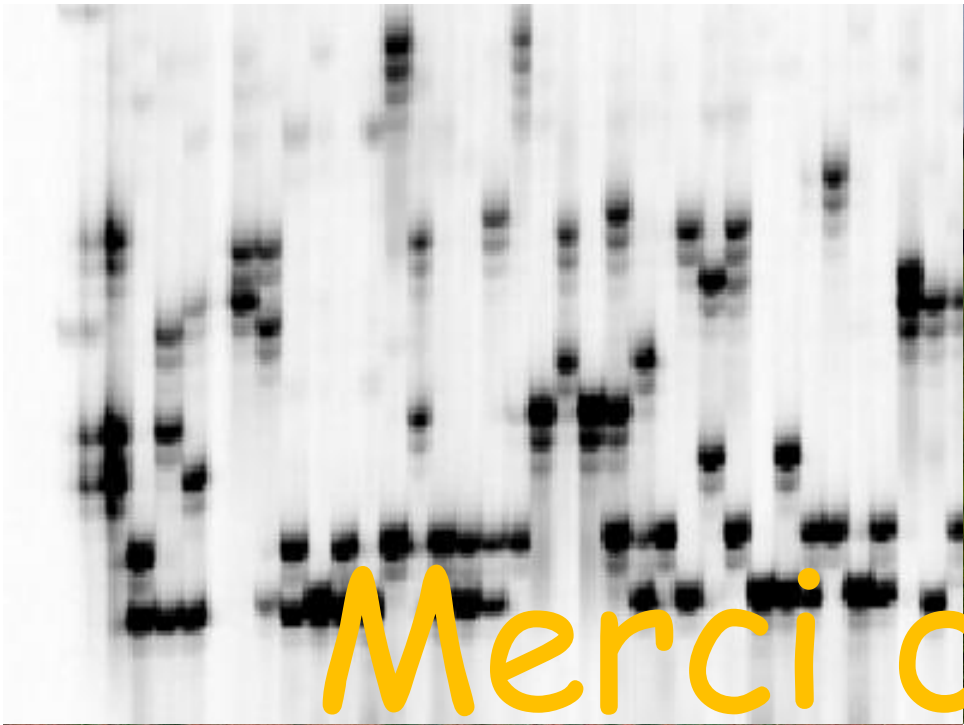
Conclusions

- Nécessité de prendre en compte le niveau d'amplification et d'augmenter la profondeur de séquençage
- Nécessité de séquencer les génotypes d'intérêt pour détecter le polymorphisme
- Difficulté d'identifier des SNP "génotypables" du fait du niveau d'homozygotie et du polymorphisme élevé chez le ray-grass anglais

Apport des nouvelles méthodes de génotypage par séquençage ?

Remerciements

- Partenaires financiers:
 - Ministère français de l'agriculture: contrat de branche
 - Association des Créateurs de Variétés Fourragères (ACVF)
 - INRA: AIP Bioressources SNP_PRAIRIES
- Les acteurs:
 - URP³F : **G. Chupeau**, D. Cadier, P. Cormenier, F. Durand, S. Delaunay, J.-F. Bourcier, M. Caillaud, J.P. Sampoux
 - INRA-UBP UMR1095 Clermont-Ferrand: J. Salse et son équipe
 - INRA-EPGV Evry: D. Brunel, MC. Le Paslier, A. Bérard, A. Chauveau, R. Bounon, M. Tchoumakov et les équipes du Genoscope & CNG:C. Cruaud, P. Wincker, C. Caloustian, M. Delépine
 - IBERS: I. Armstead
 - Univ. Aarhus: T. Asp



Merci de votre



attention!

AU249224ref... X

GGAGAG - AACAAAGAT - AAATT - AGCAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 Consensus GGAGAG - AACAAAGAT - AAATT - AGCAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 Coverage 84

GGAGAG - AACAAAGAT - AAATT - AGCAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA
 GGAGAG - AACAAAGAT - AAATT - AGTAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA

Read Mapping Settings

- Read layout
- Compactness: Packed
- Gather sequences at top
- Show sequence ends
- Show mismatches
- Packed read height: Medium
- Find Conflict
- Low coverage threshold: 8
- Find Low Coverage
- Sequence layout
- No spacing
- Numbers on sequences
- Relative to: 1
- Numbers on plus strand
- Lock top sequence
- Hide labels
- Lock labels