

# Rapport de stage

Baptiste ROUGER

L2 BCST

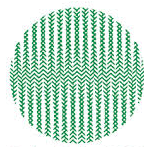
Analyse de la variation de la hauteur des  
plantes lors d'une expérience de sélection  
divergente pour la date de floraison sur le  
maïs

21 mai 2015



**Le Moulon**  
**GÉNÉTIQUE**  
**QUANTITATIVE**  
**ET ÉVOLUTION**

UMR de Génétique Quantitative et Evolution - Le Moulon  
Ferme du Moulon  
91190 Gif-sur-Yvette



Institut National de la Recherche Agronomique

**INRA**

**UNIVERSITÉ**  
**PARIS**  
**SUD**

**Directeur de recherche : Olivier MARTIN**

**Maître de stage : Christine DILLMANN**

**Enseignant référent : Judith LEGRAND**

# Table des matières

<b>Glossaire</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Le maïs, une espèce modèle d'intérêt agronomique . . . . .	3
1.1.1 Généralités sur le maïs . . . . .	3
1.1.2 Histoire du maïs . . . . .	3
1.2 Présentation de l'expérience de sélection divergente . . . . .	4
1.2.1 Généralités . . . . .	4
1.2.2 L'expérience de sélection divergente . . . . .	4
1.3 Objectifs du stage . . . . .	4
<b>2 Matériel et méthodes</b>	<b>5</b>
2.1 Le matériel végétal . . . . .	5
2.2 Méthodes . . . . .	6
2.2.1 Mise en forme des données . . . . .	6
2.2.2 Correction des données . . . . .	6
2.3 Résultats . . . . .	7
2.3.1 Réponse de la hauteur des plantes à la sélection pour la date de floraison . . . . .	7
2.3.2 Corrélation entre hauteur et date de floraison . . . . .	8
<b>3 Discussion</b>	<b>11</b>
<b>4 Conclusion</b>	<b>11</b>
<b>A Codes sources</b>	<b>13</b>

# Table des figures

1	Cycle de vie du maïs en culture . . . . .	3
2	Comparaison entre téosinte et maïs . . . . .	4
3	Réponse des deux lignées pures de maïs à la sélection pour la date de floraison	5
4	Principe de l'expérience de sélection divergente . . . . .	6
5	Généalogie des plantes sélectionnées chez les deux lignées de maïs F252 et MBS pour l'expérience de sélection divergente en 2012 . . . . .	6
6	Représentation des hauteurs des plantes de la lignée F252 . . . . .	7
7	Représentation des hauteurs des plantes de la lignée MBS . . . . .	8
8	Représentation de la relation entre hauteur et date de floraison . . . . .	8
9	Représentation séparée de la corrélation entre hauteur et date de floraison pour deux familles . . . . .	9
10	Représentation des coefficients de corrélation entre hauteur et date de flo- raison par plante en 2013 et 2014 pour F252 . . . . .	10
11	Représentation des coefficients de corrélation entre hauteur et date de flo- raison par plante en 2013 et 2014 pour MBS . . . . .	10

# Glossaire

**Degré-jour :** unité de mesure utilisée pour calculer l'accumulation de chaleur nécessaire à la durée d'un développement biologique comme la croissance d'une plante.

**Dérive génétique :** réduction du nombre de génotypes différents dans une population de taille finie du fait de l'échantillonnage aléatoire des gamètes ou des individus.

**Monoïque :** se dit d'une plante dont les fleurs mâles et femelles sont distinctes mais situées sur le même pied.

**Mutation :**

**Sélection naturelle :** pression évolutive qui conduit à des changements de la composition génétique d'une population d'une génération à l'autre, en lien avec des différences génétiques des capacités de survie ou de reproduction.

# 1 Introduction

## 1.1 Le maïs, une espèce modèle d'intérêt agronomique

### 1.1.1 Généralités sur le maïs

Le maïs (*Zea mays*) est une plante herbacée annuelle appartenant à la famille des *Poacées* (graminées). Par ailleurs, il s'agit d'une plante monoïque.

**Cycle de vie du maïs** Le cycle de vie du maïs (FIGURE 1) commence par la phase de dormance du grain, puis sa germination. Vient ensuite le stade végétatif jeune durant lequel la plante développe ses feuilles juvéniles. Il s'achève par la transition florale qui marque le début du stade végétatif adulte. Celui-ci se termine par la floraison débutant le stade reproducteur. Pour finir, lorsque les graines sont mûres, celles-ci peuvent être récoltées.

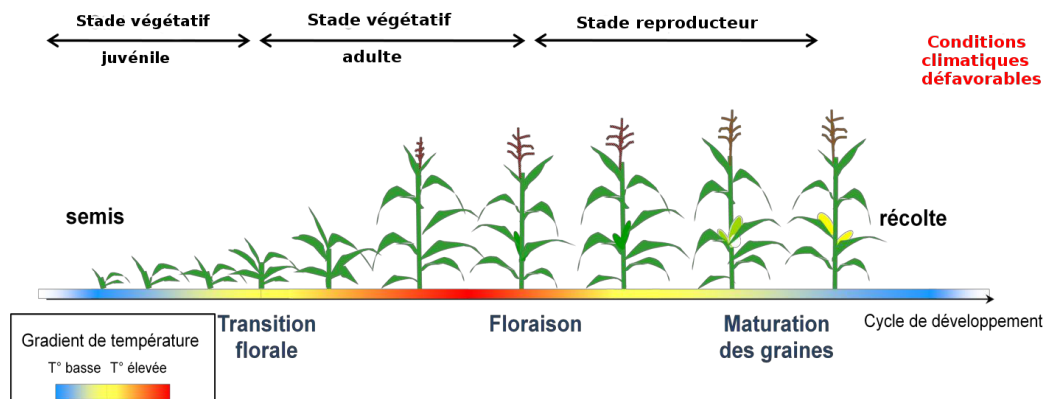


FIGURE 1 – Cycle de vie du maïs en culture

### 1.1.2 Histoire du maïs

Le maïs est une espèce domestiquée depuis environ 9000 ans. Elle est originaire des alentours de Mexico et est issue de la téosinte (*Zea sp*). Elle s'est rapidement propagée à travers le continent américain[1] en quelques milliers d'années. Cette expansion demanda son adaptation à des climats plus tempérés que son environnement naturel tropical, ce qui fut réalisé en sélectionnant les variétés insensibles à la photopériode, relativement précoces et capables d'effectuer leur cycle de croissance avant l'arrivée des conditions climatiques défavorable (FIGURE 1). Elle fut découverte et ramenée pour la première fois en Europe par Christophe Colomb lors de sa découverte de l'Amérique. C'est aujourd'hui une des premières céréales cultivées dans le monde.

**Domestication du maïs** Lors de sa domestication, le maïs a vu certaines de ses caractéristiques changer pour le voir devenir ce qu'il est aujourd'hui (FIGURE 2). On peut par exemple citer l'absence de dissémination de ses graines, l'augmentation de leur poids et de leur nombre, la synchronisation des dates de floraison mâles et femelles ainsi qu'une dominance apicale entraînant une modification de l'architecture de la plante.

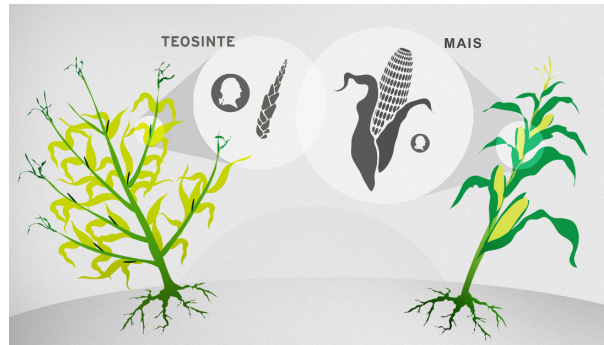


FIGURE 2 – Comparaison entre téosinte et maïs

## 1.2 Présentation de l'expérience de sélection divergente

### 1.2.1 Généralités

**Des forces évolutives** Les individus d'une population interagissent de façon dynamique avec leur environnement. Ces interactions conduisent à des changements de composition génétique des populations sous l'effet des pressions évolutives. Ainsi, les mutations représentent une pression évolutive puisqu'elles modifient la composition génétique des individus d'une population. De la même façon, on peut citer la dérive génétique et la sélection naturelle, qui amènent à la conservation de certains génotypes.

**L'adaptation** Il s'agit d'un phénomène résultant des forces évolutives, telle que la sélection. Elle témoigne de la plus grande capacité des individus adaptés à survivre dans leur environnement et à s'y reproduire.

### 1.2.2 L'expérience de sélection divergente

L'expérience de sélection divergente a débuté en 1993. Prenant pour origine deux lignées pures de maïs (F252 et MBS847) à très forte homozygotie, les plants les plus précoces et les plus tardifs pour la date de floraison furent sélectionnés pour le semis chaque année. On obtient ainsi deux populations de maïs pour une même lignée pure, marquant une plus ou moins forte réponse à la sélection selon la variété et la direction de la sélection (FIGURE 3).

Par ailleurs, la taille des plantes fut aussi mesurée, pour observer si, comment, et combien un caractère n'intervenant pas dans la sélection artificielle d'un individu pouvait être modifié par celle-ci.

## 1.3 Objectifs du stage

Une analyse fine des paramètres de taille pour chaque population de maïs doit être faite pour déterminer si, comme nous le pensons, la hauteur est corrélée à la date de floraison, et de quelle façon (causale, aléatoire...), et si c'est le cas, dans quelle mesure.

Après nous être intéressés au matériel utilisé lors des vingt années d'expérimentation, nous nous intéresserons aux méthodes de traitement des données. Nous analyserons ensuite les résultats et les interpréterons. Enfin, nous conclurons ce sujet.

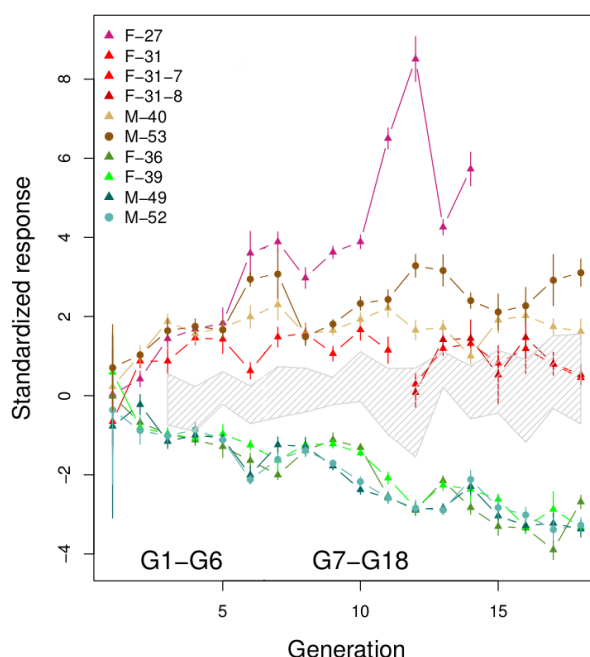


FIGURE 3 – Réponse des deux lignées pures de maïs à la sélection pour la date de floraison

## 2 Matériel et méthodes

### 2.1 Le matériel végétal

L'expérience de sélection divergente (FIGURE 4), débutée en 1993 et toujours en cours, compte 19 générations. Les lignées de départ utilisées sont deux lignées pures à forte homozygotie : les maïs F252 et MBS847 (notée MBS). On obtient pour chaque lignée deux populations, une tardive et une précoce, sélectionnées pour la date de floraison<sup>1</sup> notée en degrés-jours. On replante également chaque année la lignée pure de départ comme témoin. Par ailleurs, chaque lignée est représentée par plusieurs familles tardives et précoces (FIGURE 3) dont le nom est composé d'une lettre suivie de deux ou trois chiffres. Chaque famille est représentée chaque année par un certain nombre de descendants des individus sélectionnés à la génération précédente. Ces individus, appelés « progéniteurs » sont nommés d'une façon qui permet de retracer facilement leur généalogie (FIGURE 5 : chaque point représente un progéniteur sélectionné pour être semé. Les traits entre les points indiquent les relations de parenté entre les progéniteurs. Les derniers progéniteurs sont marqués en noir ainsi que tous leurs ancêtres pour pouvoir retracer leur généalogie.). On a donc plusieurs niveaux d'observation des variations étudiables : progéniteur, famille, population et lignée.

A chaque génération, on réalise pour chaque lignée une autofécondation<sup>2</sup> des dix individus les plus précoces et les plus tardifs.

1. Définie lorsque les fleurs mâles et femelles apparaissent

2. Pour réaliser l'autofécondation, on couvre les fleurs des individus de sachets pour recueillir le pollen et protéger les soies. Après vingt-quatre heures, le sachet contenant le pollen est déposé sur la fleur femelle pour réaliser l'autofécondation.

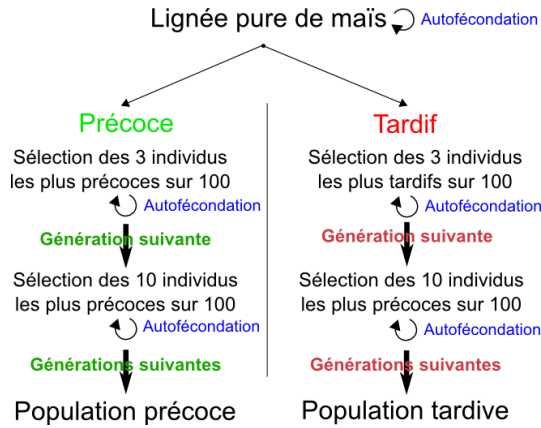


FIGURE 4 – Principe de l’expérience de sélection divergente

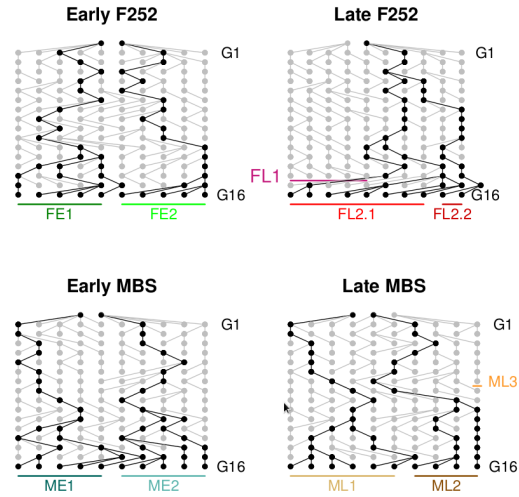


FIGURE 5 – Généalogie des plantes sélectionnées chez les deux lignées de maïs F252 et MBS pour l’expérience de sélection divergente en 2012

## 2.2 Méthodes

Pour réaliser les analyses nécessaires à la confirmation de nos hypothèses, il faut commencer par formater toutes les données au même format, c’est à dire que chaque plante mesurée sera référencée avec de nombreuses caractéristiques la concernant : l’année, la génération, la lignée (F252 ou MBS), le parent (c’est-à-dire le « grand-parent » de la plante mesurée), le progéniteur (le parent de la plante), la population à laquelle elle appartient (tardive, précoce ou témoin), la famille qui permet de retracer la généalogie, le bloc, la ligne et le numéro de la plante qui permettent de retrouver la position de la plante dans le champs, et bien évidemment la hauteur de la plante.

### 2.2.1 Mise en forme des données

L’ensemble des données que l’on doit rassembler se trouve dans des fichiers de formats différents. Il faut donc les mettre au même format pour commencer à traiter les données. On choisi le format **CSV** pour des raisons de commodité. Ces fichiers sont ensuite traités par le script<sup>3</sup> *tri.R*. Celui-ci retourne alors un fichier de données formatées par année. Puis le script *aggregate.R* permet de rassembler toutes ces données dans un même fichier au format **CSV**.

### 2.2.2 Correction des données

L’impact de l’année et de l’emplacement dans le champ (appelé « effet bloc ») sur la hauteur des plantes a ensuite été constaté . En effet, il apparaît clair que l’année va jouer un rôle important sur la hauteur de la plante, en fonction de la quantité de précipitations, de la température cumulée, etc. De la même façon, les propriétés du terrain peuvent varier localement, selon par exemple la composition du sol, ce qui va là encore influencer sur

3. Tous les scripts sont réalisés dans le langage interprété par le logiciel d’analyses statistiques **R**. cf ANNEXES A

la hauteur des plantes. Un script<sup>4</sup> a donc été réalisé pour estimer et corriger les données récoltées.

Ce script réalise une analyse (effet de l'année et « effet bloc ») de la variance de la hauteur. On utilise ensuite les résultats de cette analyse pour corriger dans un premier temps l'« effet bloc », et dans un second temps l'effet année. Ceci peut être fait grâce aux plantes témoins, qui nous permettent de supposer que leur hauteur ne varie pas et de vérifier que la correction des données est efficace.

## 2.3 Résultats

### 2.3.1 Réponse de la hauteur des plantes à la sélection pour la date de floraison

Les deux figures suivantes (FIGURES 6 et 7) représentent par des boxplots les variations de la hauteur des individus mesurés pour chaque population en fonction du temps. Elles présentent également la pente de la variation des hauteurs au cours du temps (on peut trouver le coefficient directeur de cette droite en haut de chaque graphe).

Si l'on observe les témoins, on remarque que le coefficient de leur droite de régression est de zéro (avec une certitude de rejet de  $H_0$ (variation de hauteur en fonction du temps) de 100%). On peut ainsi dire que les données ont bien été corrigées puisque les hauteurs des témoins sont stables et invariées au cours du temps.

On observe également (FIGURE 6 et FIGURE 7) des tendances à la hausse pour les populations tardives (0.53 cm/an et 0.21 cm/an respectivement pour F252 et MBS) et une tendance à la baisse pour la population précoce de MBS (-0.64cm/an). La population précoce de F252 semble cependant augmenter de taille, bien que très faiblement (0.03cm/an). D'autre part, sur les mêmes graphiques mais concernant les populations précoces et tardives des deux lignées, on peut remarquer que les variations des hauteurs des plantes au cours du temps sont très faibles pour l'ensemble des données observées. Or on constate, particulièrement pour les lignées tardives, que la hauteur des plantes est déjà différente de la hauteur du témoin pour la première année d'observation des hauteurs. Cela semble indiquer qu'une réponse très rapide à la sélection a eu lieu dans les premières années, puis que cette réponse a très rapidement diminué voire disparu.

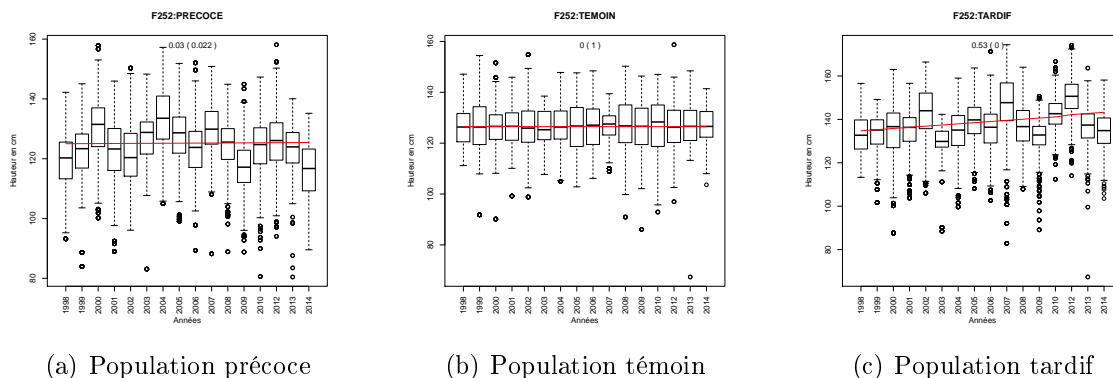


FIGURE 6 – Représentation des hauteurs des plantes de la lignée F252

---

4. Il s'agit du script *DataCorrectv2.R*



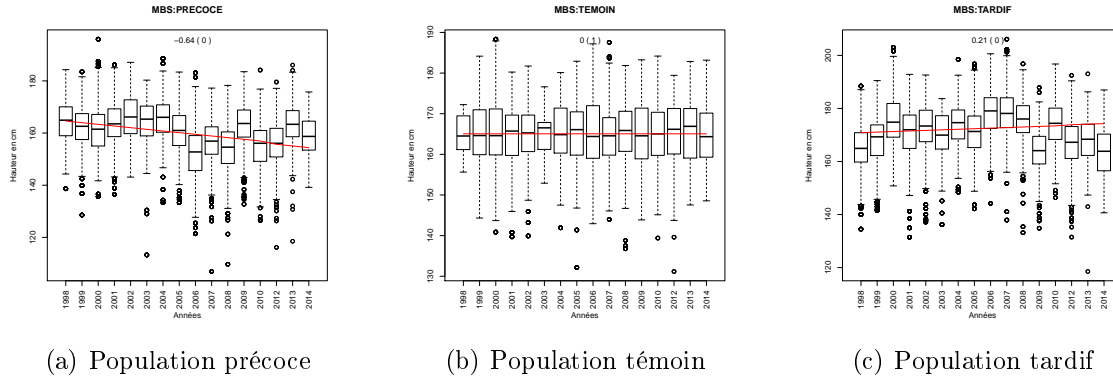


FIGURE 7 – Représentation des hauteurs des plantes de la lignée MBS

### 2.3.2 Corrélation entre hauteur et date de floraison

**Corrélation par famille et par année** Il a également été testée si il y avait bien une corrélation entre la hauteur des plantes et la date de floraison (FIGURE 8). Pour cela, on a représenté pour chaque famille et chaque année la hauteur moyenne des plantes en fonction de leur date moyenne de floraison.

On observe ainsi une répartition globale des points autour d'une droite. On peut également observer une scission entre les populations précoces et tardives des lignées (respectivement représentées dans des couleurs froides et des couleurs chaudes en FIGURE 8). Par ailleurs, il semble qu'il n'y ait pas forcément de corrélation intra-famille.

Cela confirme donc le fait que la hauteur et la date de floraison sont corrélées pour les lignées.

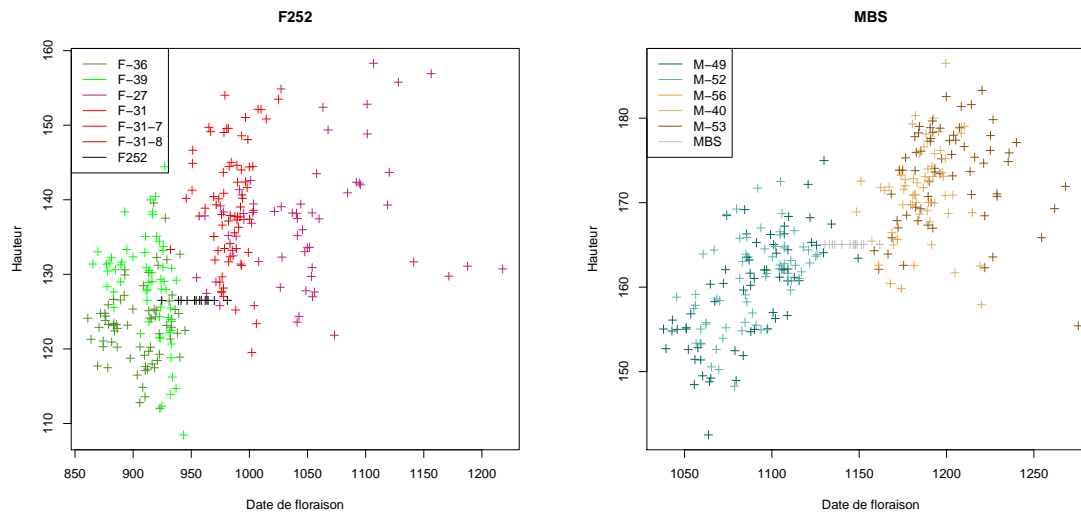


FIGURE 8 – Représentation de la relation entre hauteur et date de floraison  
Chaque point d'une couleur représente la moyenne de la hauteur des progéniteurs d'une famille pour une année (couleurs dans la figure) en fonction de la date moyenne de floraison de ces mêmes plantes. A gauche, la lignée F252 et à droite la lignée MBS.

**Corrélation intra-famille** En FIGURE 9, on a représenté pour deux familles précoces (F-39 et M-49) l'évolution de la relation entre hauteur et date de floraison en indiquant

l'année par des nombres. Ainsi, on peut suivre l'évolution de la corrélation. On observe ainsi que les points ne semblent pas évoluer dans le même sens pour la hauteur : on voit F252 augmenter puis se stabiliser, tandis que MBS voit sa hauteur descendre de façon semble-t-il constante. Il semble donc qu'il n'y ait pas de réelle corrélation entre la hauteur et la date floraison.

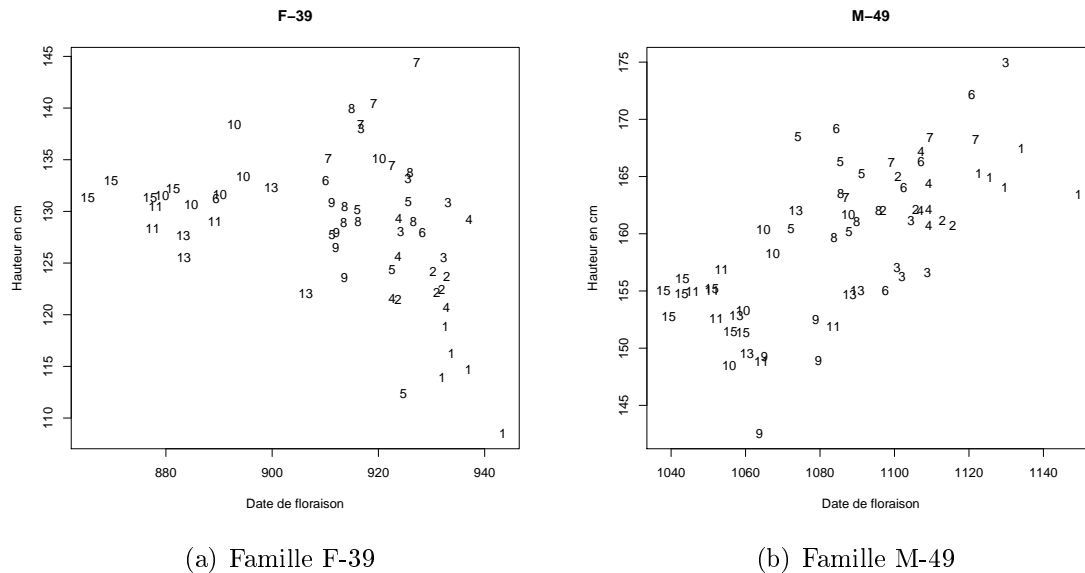


FIGURE 9 – Représentation séparée de la corrélation entre hauteur et date de floraison pour deux familles

Chaque nombre représente par sa position la moyenne de la hauteur des progéniteurs de la famille (F-39 à gauche et M-49 à droite) en fonction de la date de floraison. Les nombres indiquent l'année (1998 étant l'année 1), on peut donc suivre l'évolution de la corrélation au cours du temps.

**Corrélation par plante pour 2013 et 2014** Les FIGURES 10 et 11 présentent les coefficients de corrélation de chaque famille pour les années 2013 et 2014 pour les deux lignées. Il s'agit de données mesurée plante par plante et non corrigées. Si l'on regarde tout d'abord au niveau global (c'est à dire sur l'ensemble de la lignée de l'année), on remarque que la corrélation est toujours significative à 95% et positive. A l'échelle des populations (précoce, témoin et tardif), on peut observer des inconstances dans les corrélations. Ainsi, pour F252, on peut voir que les tardifs ont une corrélation non significative pour les deux années, alors que les populations précoce et témoin ont un coefficient de corrélation tantôt significatif, tantôt non significatif. Pour MBS, on remarque que la population précoce reste corrélée de façon significative et que la population témoin a un coefficient de corrélation qui reste non significatif. On voit également, cependant, que la population témoin est tantôt significative, tantôt non significative.

Quant aux familles, elles ont un coefficient de corrélation pour la plupart non significatif. On observe cependant des années où les familles sont corrélées de façon significative, sans lien avec les autres années.

De même, on peut observer que, pour certaines familles et populations, les coefficients de corrélation sont tantôt positifs, tantôt négatifs.

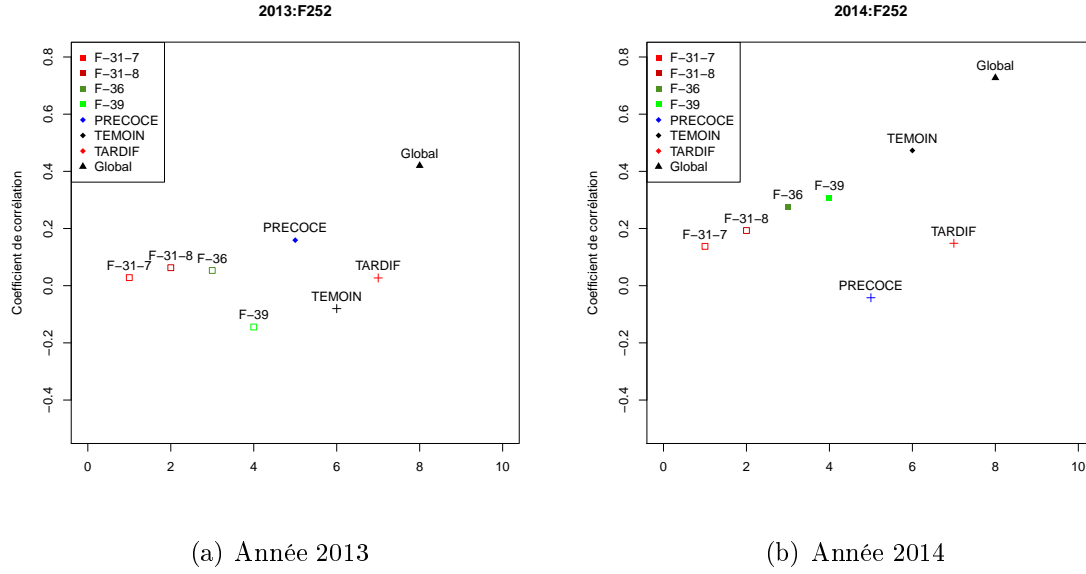


FIGURE 10 – Représentation des coefficients de corrélation entre hauteur et date de floraison par plante en 2013 et 2014 pour F252

Le coefficient de corrélation de chaque famille, calculé en associant la hauteur et la date de floraison de chaque plante, est représenté en figuré plein si la corrélation est significative à 95% et en figuré creux si la corrélation n'est pas significative pour le même seuil.

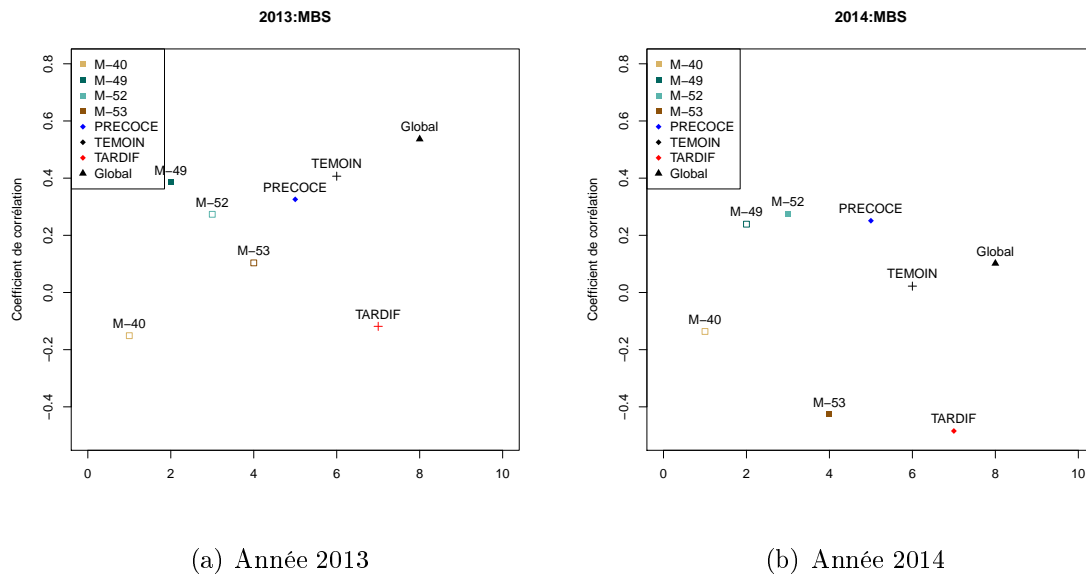


FIGURE 11 – Représentation des coefficients de corrélation entre hauteur et date de floraison par plante en 2013 et 2014 pour MBS

Le coefficient de corrélation de chaque famille, calculé en associant la hauteur et la date de floraison de chaque plante, est représenté en figuré plein si la corrélation est significative à 95% et en figuré creux si la corrélation n'est pas significative pour le même seuil.

On peut donc dire qu'il n'y a pas de corrélation intra-famille ou populationnelle entre la hauteur et la date de floraison. Cependant, sur l'ensemble des plantes, on peut confirmer l'existence d'une corrélation entre la hauteur et la date de floraison.

### 3 Discussion

On a donc pu conclure à une corrélation entre la hauteur et la date de floraison sur l'ensemble des plantes d'une lignée, et ce de façon certaine puisque celle-ci apparaît toujours de façon positive et significative, mais que cette corrélation n'existait pas avec certitude à l'intérieur des populations et des familles. Si le biais échantillonnage paraît improbable (les échantillons vont de quelques centaines d'individus à des milliers d'individus mesurés), un biais a pu être introduit lors de la correction des données, ce qui induirait une erreur dans les tests et les représentations. De la même façon, lors de l'analyse de la corrélation par plante pour 2013 et 2014, les données n'ont pas été corrigées, et un biais a pu être introduit par « effet bloc » ou « effet année ». Il pourrait donc être important de corriger ces données.

Nous avons vu qu'une corrélation entre hauteur et date de floraison existait pour les lignées. Il serait intéressant de voir si le nombre de feuille et/ou la longueur des entrenœuds a été modifiée par l'expérience de sélection divergente pour la date de floraison.

### 4 Conclusion

Nous avons donc montré qu'une corrélation, comme nous le pensions, existait entre la hauteur et la date de floraison, mais que celle-ci n'existait qu'au niveau des lignées, et pas au niveau des familles. Nous n'avons pu conclure à la source de cette corrélation.

## Références

- [1] Zerjal, Rousselet, Mhiri, Combes, Madur, Grandbastien, Charcosset, and Tenailon. Maize genetic diversity and association mapping using transposable element insertion polymorphisms. 2011.

## A Codes sources

Les codes sources utilisés pendant ce stage sont disponibles sur Github à l'adresse suivante : <https://github.com/rougerbaptiste/stageL2>.

Voici la fonction de chacun de ces codes :

**aggregate.R** : Ce script a pour fonction de rassembler les données contenues dans les fichiers de hauteurs de chaque année pour les rassembler dans un unique fichier.

**Datacorrectv2.R** : Ce script corrige les données des années et affiche un graphe (*Référence*) résumant les données.

**RprFam.R** : Ce script a pour fonction de représenter les données de hauteurs obtenues pour chaque famille (hauteur absolue et hauteur relative au témoin) et de créer un fichier .pdf contenant les deux graphes.

**RprPop.R** : Ce script a pour fonction de représenter les données de hauteurs obtenues pour chaque population (hauteur absolue et hauteur relative au témoin) et de créer un fichier .pdf contenant les deux graphes.

**tri.R** : Ce script a pour fonction de récupérer toutes les données brutes de hauteurs, de les mettre sous une forme standard et de créer un fichier contenant les données de hauteur par année.