



**HAL**  
open science

# Genetic mechanisms involved in climatic adaptation of experimentally evolving populations of bread wheat

Bénédicte Rhoné

► **To cite this version:**

Bénédicte Rhoné. Genetic mechanisms involved in climatic adaptation of experimentally evolving populations of bread wheat. Sciences of the Universe [physics]. AgroParisTech, 2008. English. NNT : 2008AGPT0018 . pastel-00003822

**HAL Id: pastel-00003822**

**<https://pastel.hal.science/pastel-00003822>**

Submitted on 1 Dec 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## **Bénédicte RHONÉ, 2008. – Étude de mécanismes génétiques impliqués dans l'adaptation climatique de populations expérimentales de blé tendre.**

UMR de Génétique Végétale, INRA / Univ Paris-Sud / CNRS / AgroParisTech, Ferme du Moulon, F-91190 Gif-Sur-Yvette.

---

### **RÉSUMÉ :**

L'adaptation locale résultant de pressions de sélection agissant sur des caractères héréditaires est une force majeure façonnant la diversité phénotypique des populations dans la nature. A l'heure actuelle, de nombreuses questions subsistent quant à l'impact de la sélection sur l'évolution des gènes impliqués dans l'architecture des caractères adaptatifs qui sont souvent des caractères complexes impliquant de nombreux gènes en interaction.

L'objectif de cette thèse est d'identifier certains mécanismes génétiques mis en jeu lors de l'adaptation de populations expérimentales de blé tendre à différents contextes climatiques. Le caractère étudié est la précocité de floraison, caractère adaptatif majeur chez les plantes annuelles qui permet de synchroniser l'initiation de la reproduction avec la saison favorable. Les populations expérimentales considérées dans cette étude ont une origine génétique commune et évoluent de façon indépendante depuis 1984 dans différents sites en France sans sélection humaine consciente ni migration, mais sous l'influence de la sélection naturelle et de la dérive. En se basant sur les connaissances des gènes impliqués dans la floraison chez *Arabidopsis thaliana* ou le riz et les récentes publications chez le blé, l'évolution de polymorphismes nucléotidiques de différents gènes candidats a été suivie durant 12 générations (2, 7 et 12) dans 3 populations expérimentales (Vervins au Nord de la France, Le Moulon en région parisienne et Toulouse au Sud). Ces populations ont également été caractérisées pour la précocité mesurée dans différentes conditions de longueur du jour et de vernalisation (durée d'exposition variable aux basses températures) et pour la diversité de locus microsatellites répartis sur l'ensemble du génome. La comparaison des niveaux de différenciation génétique inter population obtenue pour les trois types de diversité montre que la précocité est sélectionnée de façon divergente dès les premières générations. Cependant cette sélection ne se fait pas de façon directe et semble être le fait de la sélection agissant sur des caractères corrélés dans les différents environnements, tels que la hauteur ou le poids de grains. La sélection pour la précocité peut être mise en relation avec les caractéristiques climatiques des sites de culture, les populations du Nord fleurissant plus tardivement que la population du Sud. Cette évolution rapide du caractère s'accompagne de changements importants de fréquences alléliques pour des gènes majeurs de précocité, et la mise en place de combinaisons alléliques multi-locus spécifiques des différents contextes climatiques considérés. Cette évolution ne remet pas en cause la diversité existant à l'intérieure de chaque population puisque plusieurs combinaisons différentes sont favorisées dans les différentes populations. Ce travail de thèse ouvre des perspectives intéressantes quant au problème du maintien du potentiel adaptatif des espèces cultivées.

---

Mots-clés : Adaptation, populations expérimentales, précocité de floraison, sélection, différenciation, *Triticum aestivum* L.

## **Bénédicte RHONÉ, 2008. – Genetic mechanisms involved in climatic adaptation of experimentally evolving populations of bread wheat.**

UMR de Génétique Végétale, INRA / Univ Paris-Sud / CNRS / AgroParisTech, Ferme du Moulon, F-91190 Gif-Sur-Yvette.

---

### **Abstract :**

Local adaptation resulting from selection pressure acts on heritable characters and is a major force shaping the phenotypic diversity of natural populations. Currently, numerous questions exist surrounding the impact of selection on the evolution of genes involved in the structure of adaptive traits, which are often complex characters involving many interacting genes.

The objective of this thesis is to identify certain genetic mechanisms which are activated during the adaptation of experimental wheat populations to different climatic contexts. The adaptive trait studied was flowering time, which permits the plant to synchronize reproduction with favorable seasonal conditions. The experimental populations considered had a common genetic origin and have evolved independently since 1984 in different sites in France, without conscious human selection or migration, but subject to the influence of natural selection and genetic drift. Using a base of knowledge on genes implicated in flowering in *Arabidopsis thaliana* and rice, and recent publications on wheat, the evolution of nucleotide polymorphisms in candidate genes was followed during 12 generations (2, 7 and 12) in three experimental populations (Vervins in the north of France, Le Moulon in the Parisian region and Toulouse in the South). These populations were also characterized for precocity of flowering time measured under different day length and vernalization regimes (duration of exposure to low temperatures) and for diversity of microsatellite loci sampled from throughout the genome. The comparison of levels of genetic differentiation among populations for these three measures of genetic diversity showed that precocity was subject to divergent selection from the first generations. However, it appears that this selection was not directly for flowering time, but instead achieved through indirect selection on correlated traits in the different environments, such as plant height or grain weight. Selection for precocity can be related to climatic characteristics of the experimental locations, with the northern populations flowering later than the southern population. This rapid evolution of the trait accompanied significant changes in allele frequencies for major genes involved in flowering time, and the development of specific multi-locus allelic combinations in the different climatic contexts considered. This evolution did not diminish the existing within-population genetic diversity because multiple combinations were favored in the different populations. The work presented in this thesis is an important contribution to our understanding of the maintenance of evolutionary potential in cultivated species.

---

*Keywords:* Adaptation, experimental populations, flowering time, selection, differentiation, *Triticum aestivum* L.

## **REMERCIEMENTS**

---

A l'issu de ce travail, je tiens à remercier vivement les personnes qui ont été impliquées dans cette thèse et plus particulièrement :

**Joël Cuguen** et **Jérôme Goudet** qui ont accepté d'être rapporteurs de ce travail ainsi que **André Gallais**, **Valérie Le Corre** et **Nathalie Machon** pour leur participation au jury.

Le **Conseil Régional d'Ile de France** qui m'a accordé le financement pour mener à bien ce travail ainsi qu'**André Gallais** pour la confiance qu'il m'a témoignée en appuyant ma candidature pour la bourse auprès de l'Ecole Doctorale.

Les deux "Isa" qui ont encadré chacune à leur manière ce travail de thèse. **Isabelle Bonnin**, à l'origine du projet de thèse, m'a proposé des méthodes et des orientations de travail tout en me laissant une grande liberté d'action. Elle a veillé à tous moments à ce que je sois entourée d'une certaine aide technique pour les différentes expérimentations, ce qui m'a permis de mener à bien ce travail. **Isabelle Goldringer**, pilier de la gestion dynamique, s'est toujours montré disponible et a supporté mes questionnements intempestifs à tous moment de la journée.

L'ensemble des personnes qui m'ont aidé pour l'acquisition des nombreuses données moléculaires et phénotypiques :

**Carine Remoué**, mon "binôme" de laboratoire en début de thèse, m'a révélé ses combines de biologie moléculaire et a été un grand soutien au cours de ce travail. **Samila Mezi** a pris le relais avec enthousiasme et efficacité. L'équipe BMBI de **Mathieu Falque** m'a également apporté une précieuse aide technique : au commencement, **Xavier Raffoux** a entrepris l'extraction d'ADN, **Céline Ridel** a poursuivi avec du génotypage SNP. Au sein de l'équipe GQMS, **Delphine Madur** m'a accompagnée dans le génotypage de SNP à la **plate-forme de génotypage de l'INRA d'Evry**.

**Nathalie Galic** a été d'une aide précieuse pour toutes les expérimentations phénotypiques menées au cours de cette thèse. Merci particulièrement pour le lourd travail abattu l'Eté des récoltes en 2005, alors que je m'étais évadée à l'autre bout du

monde... Je remercie également **Nadège Tuo** dont la présence s'est révélée indispensable au moment des notations d'épiaisons, les nombreux mains-d'œuvre qui ont participé au travail de terrain et aux longues séances de battage et Didier Tropée pour la maintenance de la capricieuse chambre de vernalisation.

**Jacques Legouis** et **Michel Rousset** sont à l'origine de l'expérimentation en chambre photopériodique à l'INRA de Mons. **Emmanuel Heumez** a pris en main l'intégralité de la manip avec beaucoup de dévouement, secondé par **Jean-Pierre**.

La **plateforme de génotypage de l'INRA de Clermont-Ferrand**, dirigé par **Gilles Boutet** à pris en charge le génotypage des marqueurs microsatellites.

La **Société Biogemma** et **Sébastien Praud** en particulier m'ont donné accès à leur connaissance du gène *Ppd* et m'ont accueilli m'accueilli dans le laboratoire pour le génotypage.

Je remercie également:

**Renaud Vitalis** pour son aide lors de l'analyse des données, en particulier pour sa grande disponibilité au cours de l'Été 2007 quelques semaines avant l'ESEB alors que j'étais en panique !

**Joëlle Ronfort**, **Frédérique Viard**, **Jacques David**, **Christine Dillman** qui ont participé à mes comités de thèse.

**Françoise Launay** et **Corinne Fiers** de l'Ecole Doctorale ABIÉS pour leur soutien logistique et leur grande efficacité.

Les **Isa**, **Yves Rousselle**, **Carine Remoué** et **Pierre Alexis** qui ont pris le temps de relire tout ou parties de ce document.

Je remercie enfin **ma famille** et **Pierre-Alexis** pour leur soutien et leurs encouragements.

# SOMMAIRE

---

<b>CHAPITRE I : INTRODUCTION</b>	<b>4</b>
<b>I-1 L'adaptation comme moteur de l'évolution</b>	<b>4</b>
I-1.1 Définitions et exemples	4
I-1.2 Intérêt d'étudier l'adaptation	5
<b>I-2 Comprendre les mécanismes génétiques de l'adaptation</b>	<b>6</b>
I-2.1 Dissection des bases génétiques de l'adaptation	6
I-2.1.1 Architecture des caractères quantitatifs adaptatifs	7
I-2.1.2 Détection de la sélection	9
I-2.1.3 Combinaisons de plusieurs informations	11
I-2.2 Des avancées, mais encore de nombreuses questions	11
I-2.2.1 Nombre et effets des gènes impliqués dans l'adaptation	12
I-2.2.2 « Standing genetic variation » ou nouvelles mutations ?	13
I-2.2.3 Rôle de la pléiotropie	15
I-2.2.4 Rôle de l'épistasie	15
I-2.2.5 Type des mutations préférentiellement sélectionnées	16
I-2.2.6 Conclusions	16
<b>I-3 Etude de l'adaptation locale dans des populations expérimentales de blé hexaploïde : exemple de la précocité de floraison</b>	<b>17</b>
I-3.1 Présentation des populations expérimentales de gestion dynamique	17
I-3.1.1 Principe des populations	17
I-3.1.2 Principaux résultats	18
I-3.3 Pourquoi s'intéresser à la précocité de floraison en particulier ?	20
I-3.3.1 La précocité de floraison, un caractère adaptatif majeur.	20
I-3.3.2 L'architecture du caractère chez le blé	21
I-3.3.3 Déterminisme bien connu pour des espèces modèles tels le riz ou <i>Arabidopsis thaliana</i>	21
<b>I-4 Objectifs et plan de la thèse</b>	<b>24</b>
<b>CHAPITRE II : MATERIEL ET METHODES</b>	<b>26</b>
<b>II-1. Matériel végétal</b>	<b>26</b>
II-1.1 Conduite des populations de gestion dynamique	26
II-1.2 Echantillonnage	27
II-1.3 Constitution du matériel d'étude	27
<b>II-2. Dispositif expérimental pour la caractérisation phénotypique des populations</b>	<b>28</b>
II-2.1 Estimation de la précocité de floraison au champ et sous différentes conditions de vernalisation en pépinière	28
II-2.2 Estimation des composantes de la précocité à partir des différents traitements	32
II-2.3 Estimation des composantes de la valeur sélective	33
<b>II-3. Marquage microsatellite des populations de GD</b>	<b>34</b>
II-3.1 Extraction	34
II-3.2 Choix des marqueurs	34
II-3.3 Génotypage et acquisition des résultats	36
<b>II-4 Génotypage des gènes candidats</b>	<b>36</b>
II-4.1 Révélation du polymorphisme de <i>VRN-1</i>	36
II-4.2 Révélation du polymorphisme de <i>Ppd-1</i>	37
II-4.3 Génotypage de <i>FT</i>	38
II-4.4 Génotypage de <i>LD</i> , <i>CO</i> et <i>GI</i>	38

## **CHAPITRE III : SUIVI DE L'EVOLUTION DES MARQUEURS NEUTRES DU GENOME DANS LES POPULATIONS DE GESTION DYNAMIQUE 41**

### **III-1 Introduction 41**

### **III-2. Méthodes 45**

- III-2.1 Diversité 45
- III-2.2 Taux d'allo-fécondation 45
- III-2.3 Structuration spatiale et temporelle 47
- III-2.4 Variations de fréquences alléliques et effectifs efficaces 47
- III-2.5 Déséquilibre de liaison 49
- III-2.6 Reconstruction d'haplotypes multilocus 50

### **III-3 Résultats 52**

- III-3.1 Diversité génétique, apparition de nouveaux allèles 52
- III-3.2 Différenciation spatiale 54
- III-3.3 Variations temporelles des fréquences alléliques et effectifs efficaces 55
- III-3.4 Régime de reproduction 56
- III-3.5 Evolution du DL et génotype multilocus 57

### **III-4. Discussion 65**

- III-4.1 Régime de reproduction 65
- III-4.2 Régime de reproduction et effectif efficace 68
- III-4.3 Régime de reproduction et déséquilibre de liaison 69
- III-4.4 Nouveaux allèles : mutation ou migration ? 71
- III-4.5 Cas particulier de la population de Vervins 73

### **III-5 Conclusions 75**

## **CHAPITRE IV: EVOLUTION DES CARACTERES QUANTITATIFS: PRECOCITE DE FLORAISON ET COMPOSANTES DE LA VALEUR SELECTIVE 77**

### **IV-1 Introduction 77**

### **IV-2 Méthodes 80**

- IV-2.1 Estimation des variances environnementales 81
- IV-2.2 Evolution des caractères au cours des générations dans chaque population 84
  - IV-2.2.1 Evolution phénotypique des caractères 84
  - IV-2.2.2 Caractérisation de la variabilité génétique intra population des caractères 85
- IV-2.3 Caractérisation de la variabilité génétique inter populations des caractères : mise en évidence de la sélection. 86
- IV-2.4 Estimation de la sélection directionnelle directe et indirecte : approche multivariée 87
  - IV-2.4.1 Estimation des gradients de sélection directionnelle observés dans les conditions "champ" au Moulon. 87
  - IV-2.4.2 Estimation des gradients de sélection attendus dans des contextes variant pour les besoins en vernalisation 89

### **IV-3. Résultats 91**

- IV-3.1 Evolution phénotypique de la précocité et des composantes de la valeur sélective au champ 91
- IV-3.2 Evolution de la variabilité intra population pour les caractères mesurés au champ 93
- IV-3.3 Détection de la sélection 95
- IV-3.4 Corrélations entre précocité et composantes de la valeur sélective : estimation des gradients de sélection. 96
- IV-3.5 Estimation des gradients de sélection attendus en conditions différentes pour la durée de vernalisation. 97
- IV-3.6. Evolution des composantes de la précocité de floraison 99
  - IV-3.6.1 Evolution des composantes de la précocité dans les populations 99
  - IV-3.6.2 Evolution des corrélations entre mesures de précocité 102

# *Chapitre 1*

## *INTRODUCTION*

# Chapitre I : Introduction

## I-1 L'adaptation comme moteur de l'évolution

### I-1.1 Définitions et exemples

L'adaptation est le processus selon lequel une population accroît sa survie et son succès reproducteur dans un environnement donné par l'action de la sélection naturelle. Celle-ci peut agir si les individus de la population ne sont pas tous identiques et qu'une part de cette variabilité affecte des caractères jouant sur la valeur sélective qui est définie comme le produit de la survie et du succès reproducteur (nombre de descendants viables). L'évolution d'une population par l'action de la sélection nécessite de plus que la variabilité de ces caractères soit contrôlée génétiquement, au moins partiellement, et ne varie pas uniquement en fonction de l'environnement, de façon à ce que les changements phénotypiques soient transmis à la génération suivante. La sélection agit donc sur la part héritable des caractères dits adaptatifs, c'est-à-dire corrélés avec le succès reproducteur, pour produire des changements adaptatifs.

Par exemple chez les plantes annuelles, la date de floraison est un caractère adaptatif important. En effet, le passage de l'état végétatif à l'état reproducteur est un événement majeur du développement souvent irréversible ; pour que le succès reproducteur soit assuré cette transition doit être initiée lorsque les conditions abiotiques (température, disponibilité des nutriments, photopériode) et biotiques (pressions parasitaires et herbivores, présence des pollinisateurs) du milieu sont favorables. Les variations de la date de floraison en fonction des conditions du milieu ont pu être mises en évidence dans la nature chez de nombreuses espèces (Loskutov 2001 ; Boudry *et al.* 2002 ; Sandring *et al.* 2007). Ces études montrent que la date de floraison est un caractère fortement héritable et que les différences de date de floraison dans divers environnements sont le fait de pressions de sélection qui s'exercent sur le caractère et conduisent à l'adaptation des populations à leur environnement.

Cependant, les caractères hérifiables peuvent également évoluer de façon aléatoire : tous les individus viables d'une population n'auront pas nécessairement de descendants à cause d'aléas environnementaux ou de la taille limitée de la population qui impose alors un

échantillonnage des individus. Cet échantillonnage peut notamment dépendre du régime de reproduction et est d'autant plus important que la taille de la population qui se reproduit (taille effective ou efficace) est limitée. On parle alors d'évolution neutre des caractères et de dérive génétique (les gènes qui contrôlent ces caractères dérivent aléatoirement). Cet effet peut être déterminant en particulier pour les caractères hérissables peu corrélés à la valeur sélective qui sont alors « neutres » vis-à-vis de la sélection. La dérive génétique est d'autant plus forte que l'effectif de la population est faible. Ainsi pour comprendre l'évolution d'un caractère, il est nécessaire de déterminer quelle part de cette évolution est due à la dérive aléatoire et quelle part est due à la sélection naturelle, c'est-à-dire à l'adaptation proprement dite.

### **I-1.2 Intérêt d'étudier l'adaptation**

L'adaptation de populations évoluant dans des environnements contrastés participe au maintien de la diversité dans la nature. En effet, s'il existe une variabilité génétique suffisante (qui apparaît au hasard par les mutations), la majorité des organismes vivants sont alors capables d'évoluer dans une grande variété de milieux. Dans chaque environnement, si la sélection naturelle est efficace (peu de dérive), le phénotype moyen de la population tend à évoluer vers le phénotype présentant la meilleure valeur sélective dans cet environnement. Ainsi, l'adaptation locale des populations à chaque environnement particulier (ou micro-milieu) maintient une diversité phénotypique au sein de l'espèce. Cette diversité participe à sa survie à l'échelle de l'écosystème puisqu'elle est disponible pour une nouvelle adaptation en cas de perturbation du milieu. A l'extrême, si au cours de l'adaptation des barrières à la reproduction se mettent en place du fait de l'isolement, la divergence prolongée des populations peut finalement conduire à l'apparition de nouvelles espèces. Le débat sur la relation entre macroévolution (*i.e.* évolution des espèces) et microévolution (*i.e.* évolution des populations) n'est pas clos, mais il semble désormais clair que les processus qui déterminent la microévolution, tels que la mutation, la dérive génétique et la sélection, permettent d'expliquer les grandes tendances observées d'un point de vue macroévolutif (Hendry et Kinnison 2001).

Du fait de la prise de conscience récente de l'impact de l'activité humaine sur l'évolution de la diversité des espèces sauvages et cultivées (exploitation, dégradation des habitats et modification du climat), la biologie de la conservation a connu un essor important ces dernières années (Stockwell *et al.* 2003). Il en ressort finalement que l'un des enjeux majeurs

de cette discipline est le maintien de la variabilité adaptative. En effet la capacité d'une population à s'adapter aux changements futurs, qu'ils soient liés à l'environnement (modification du milieu, changement climatique global, évolution des populations de parasites), à l'évolution de la demande de la société (agriculture plus respectueuse de l'environnement, qualité des produits, diversification) ou à la demande des producteurs (maintien du revenu dans un contexte de baisse des prix de vente, limitation du nombre d'intermédiaires) est déterminée par la disponibilité de la variabilité adaptative.

## **I-2 Comprendre les mécanismes génétiques de l'adaptation**

Au cours de l'adaptation, la sélection, qui agit directement sur le phénotype, entraîne à chaque génération des changements dans la composition génétique sous-jacente. Un phénotype est l'expression de la relation complexe entre un génotype et son environnement. Pour comprendre comment agit la sélection au cours de l'adaptation, il est tout d'abord important de pouvoir relier les caractères adaptatifs à leur base génétique puis de comprendre comment cette base génétique participe en retour à l'évolution adaptative des populations dans un environnement donné.

### **I-2.1 Dissection des bases génétiques de l'adaptation**

D'un point de vue pratique, l'étude des bases génétiques de l'adaptation reposent majoritairement sur deux types d'approches : les approches visant à comprendre l'architecture des caractères quantitatifs adaptatifs et les approches visant à détecter des régions du génome sélectionnées dans des populations naturelles. De nombreux articles ont récemment passé en revue les différentes méthodes utilisées (Vasemägi et Primmer 2005 ; Ehrenreich et Purugganan 2006 ; Mitchell-Olds *et al.* 2007 ; Stinchcombe et Hoekstra 2008), dont voici un bref aperçu.

### **I-2.1.1 Architecture des caractères quantitatifs adaptatifs**

Les études visant à identifier le déterminisme des caractères adaptatifs se basent généralement sur la détection d'associations statistiques entre les variations des caractères et celles des marqueurs génétiques dans la population. Des marqueurs génétiques (par exemple marqueurs microsatellites, Amplified Fragment Length Polymorphism ou AFLP, Single Nucleotide Polymorphism ou SNP) sont génotypés chez des individus également phénotypés pour le caractère d'intérêt. Parmi les marqueurs utilisés, les marqueurs génétiquement proches des gènes impliqués dans la variation de ce caractère sont en déséquilibre de liaison (association statistique entre allèles de locus différents, noté DL par la suite) avec le gène entraînant une association statistique entre le polymorphisme du marqueur et la variation du caractère. Ce type d'approches a historiquement été développé pour caractériser le déterminisme génétique de caractères quantitatifs des espèces d'intérêt agronomique ou chez des espèces modèles pour lesquelles beaucoup de marqueurs sont disponibles (Alonso-Blanco *et al.* 2005). Leur puissance de détection repose sur l'utilisation de nombreux marqueurs génétiques couvrant au mieux l'ensemble du génome et d'un grand nombre d'individus phénotypés. On distingue deux approches principales en fonction du type de matériel caractérisé : la cartographie de locus impliqués dans la variation d'un caractère quantitatif (QTL) et la cartographie par déséquilibre de liaison (LD mapping) ou étude d'association.

#### ***I-2.1.1.1 La cartographie de locus impliqués dans la variation d'un caractère quantitatif ou QTL (Quantitative Trait Loci)***

La recherche d'association entre les marqueurs génétiques et le phénotype par détection de QTL se base sur l'utilisation de croisements contrôlés d'individus différenciés pour le caractère d'intérêt afin de détecter les régions du génome expliquant cette variabilité phénotypique. Dans le cas général, deux lignées parentales présentant des phénotypes contrastés pour le caractère étudié sont croisées pour former une population F1. Puis les individus de cette population F1 sont à leur tour croisés entre eux pour former une population F2 ou sont rétro-croisés avec l'un des deux parents. La descendance F2 ainsi obtenue présente une distribution continue, pour un caractère polygénique, car l'ensemble des gènes parentaux gouvernant ce caractère sont mixés aléatoirement par recombinaison et donnent lieu à de nouvelles combinaisons alléliques d'effets quantitatifs différents. Dans le cas d'individus

allogames, des schémas plus complexes de croisements doivent cependant être envisagés (cf. Lynch et Walsh 1998 ou Slate 2005). Les marqueurs sont alors ordonnés entre eux pour construire une carte génétique puis les locus significativement corrélés à la variation du caractère sont identifiés. Cette approche permet d'estimer le nombre minimal de locus impliqués dans la variation d'un caractère (*i.e.* la part de variance génétique expliquée par la variation au QTL) pour un croisement et un environnement donnés ainsi que leur position approximative et leur effet sur le caractère. Notons que cette approche ne permet pas d'identifier directement les gènes mais seulement des segments chromosomiques compris entre 2 marqueurs et qui peuvent contenir plusieurs gènes, la précision maximale observée jusqu'ici étant de 2 cM (Price 2006). Par ailleurs, le nombre de QTL détectés dépend du polymorphisme des parents utilisés pour le croisement. L'utilisation de plusieurs croisements de parents différents et de plusieurs environnements permet de mieux appréhender l'architecture du caractère.

#### ***I-2.1.1.2 La cartographie par déséquilibre de liaison (LD mapping) ou étude d'association***

Dans le cas de la cartographie par déséquilibre de liaison, l'association entre les marqueurs moléculaires et le phénotype est testée à partir des données de génotypage et de phénotypage obtenue sur des collections aux bases génétiques larges. Cette approche utilise la variation naturelle de populations non apparentées ayant subi un grand nombre d'évènements de recombinaison de sorte que seuls les marqueurs moléculaires génétiquement proches des gènes impliqués dans le caractère sont corrélés avec le phénotype. La résolution de cette approche est directement liée à la décroissance du DL avec la distance physique qui varie de façon importante selon l'espèce considérée, notamment en fonction de son régime de reproduction. A titre de comparaison, chez l'espèce autogame *Arabidopsis thaliana*, la décroissance du DL a été évaluée à 250 kbp (Nordborg *et al.* 2002) alors que la distance de décroissance du DL trouvée chez le maïs à régime de reproduction allogame est de 100 à 200 paires de bases (Tenailon *et al.* 2001). Cette approche nécessite un plus grand nombre de marqueurs que la détection classique de QTL, surtout si la décroissance du DL en fonction de la distance physique est rapide, mais elle est en général plus précise. De plus, la diversité allélique considérée est plus importante que dans le cas de la détection de QTL et représente mieux la diversité génétique de l'espèce.

La détection des polymorphismes responsables de la variation du caractère est également possible par une approche de génétique d'association de polymorphisme de gènes candidats. En combinant les résultats de détection des QTL à une recherche de gènes de fonction connue et dont la séquence est disponible chez une espèce modèle apparentée, on augmente les chances de trouver une relation entre les polymorphismes de ces gènes candidats et la variation du caractère d'intérêt.

Cependant les méthodes de génétiques d'association utilisant des collections à bases génétique large peuvent conduire à l'identification de régions faussement impliquées dans la variation d'un caractère si le déséquilibre de liaison à longue distance, résultant d'une éventuelle structuration génétique des collections utilisées, n'est pas pris en compte lors de l'association marqueurs/phénotype. Des méthodes statistiques ont récemment été développées pour détecter la structuration cryptique des populations utilisées (Pritchard et *al.* 2000, Veyriéras 2006), rendant l'utilisation de ces approches plus accessible.

### **I-2.1.2 Détection de la sélection**

La plupart des méthodes visant à détecter des régions du génome ou de gènes sous sélection reposent sur le fait que les effets démographiques (goulot d'étranglement, migration, fondation) et le régime de reproduction affectent tout le génome de la même façon tandis que la sélection a une action locus-spécifique et n'affecte que quelques gènes en même temps. La recherche de signatures de la sélection dans des séquences de gènes ne sera pas considérée car ce type d'approches concerne une échelle évolutive beaucoup plus grande que celle qui est étudiée ici.

Ces approches sont basées sur le phénomène de balayage sélectif : la sélection directionnelle à un locus va entraîner l'augmentation en fréquence de l'allèle favorable ainsi que des allèles des locus avoisinants qui lui sont liés (par un effet « d'auto-stop génétique »), entraînant une perte de diversité dans son voisinage (Maynard Smith et Haigh 1974). Le génotypage d'un très grand nombre de locus indépendants, répartis sur l'ensemble du génome (d'où la dénomination « genome scan » donnée à cette approche) d'un grand nombre d'individus, issus d'une ou de plusieurs populations, permet d'identifier les régions du génome qui présentent des patrons de variations particuliers dus à la sélection. Ainsi, les régions du génome entraînées par la sélection d'un locus sont susceptibles de présenter une baisse de

diversité et un fort déséquilibre de liaison par rapport au reste du génome mais aussi un niveau de divergence élevé entre populations (Vasemägi et Primmer 2005).

(i) Détection des régions présentant une réduction de variabilité à l'intérieur d'une population

Cette méthode consiste à comparer la diversité génétique de deux échantillons en calculant la variance  $V$  des tailles d'allèles (Schlötterer 2002) ou l'hétérozygotie attendue  $H$  (Kauer *et al.* 2002) à partir de locus très polymorphes tels que les marqueurs microsatellites. En l'absence de sélection et dans certaines conditions démographiques, le ratio des variances de chaque échantillon (ou de leurs hétérozygoties) suit une distribution log-normale. Les ratios  $\ln RV$  et  $\ln RH$  sont donc calculés pour tous les locus considérés, distribués sur l'ensemble du génome, et utilisés pour construire un histogramme de leur distribution. Le ratio correspondant au locus d'une région qui a subi un important balayage sélectif dans un seul des deux échantillons sera situé à l'extrémité de la distribution avec une grande probabilité.

(ii) Détection des régions présentant un fort niveau de divergence entre plusieurs populations

Des méthodes statistiques basées sur la différenciation entre populations distinctes (mesurée par le paramètre de différenciation  $F_{ST}$ ) ont été développées pour détecter les marqueurs présentant une plus forte ou plus faible différenciation qu'attendue sous l'hypothèse de neutralité. Dans ces méthodes, les distributions attendues des mesures de différenciation sous l'hypothèse neutre sont obtenues par simulation de différents scénarios démographiques (modèle en îles, Beaumont et Nichols 1996, Beaumont et Balding 2004 ; ou modèle de populations en divergence pure, Vitalis *et al.* 2001) en se basant sur les paramètres populationnels estimés à partir de l'ensemble du jeu de données. Pour contourner le problème lié à l'utilisation de méthodes reposant sur des hypothèses de modèles évolutifs stricts, on peut utiliser plusieurs méthodes de détection de sélection sur un même jeu de données. Si le nombre de marqueurs génotypés dans les populations considérées est important, les locus sortant de la distribution peuvent être détectés à partir de la distribution empirique.

Une méthode similaire basée sur les changements temporels des fréquences alléliques (estimés par la variance standardisée des fréquences alléliques  $F_C$ ) permet d'identifier des locus présentant des valeurs extrêmes de  $F_C$  en comparaison du reste du génome (Goldringer et Bataillon 2004).

Ces méthodes, fondées sur l'étude de la diversité ou de la différenciation d'un grand nombre de locus, permettent d'identifier des régions potentiellement impliquées dans l'adaptation de populations sans qu'il ne soit nécessaire de phénotyper les individus ni de connaître le contexte écologique d'évolution ou l'agent de sélection. Elles sont donc particulièrement intéressantes lors de l'étude d'espèces non modèles pour lesquelles beaucoup d'informations font encore défaut. En revanche, le risque de détection de faux positifs est important surtout lorsqu'on manipule un grand jeu de données (Storz 2005).

### **I-2.1.3 Combinaisons de plusieurs informations**

Les deux types de méthodes présentés précédemment constituent des approches bien distinctes pour appréhender les bases génétiques de l'adaptation. L'une, en amont, vise à identifier les gènes participant à l'architecture du caractère adaptatif, l'autre, plus en aval, vise à identifier les gènes répondant à la sélection au cours de l'adaptation. Pour préciser l'action de la sélection sur les gènes impliqués dans la variation des caractères adaptatifs, certains auteurs (Vasemägi et Primmer 2005 ; Stinchcombe et Hoekstra 2008) proposent de combiner l'utilisation de ces deux types d'approches. Ainsi la co-localisation de marqueurs génétiques présentant des signatures de sélection avec des régions impliquées dans la variation du caractère adaptatif, détectées par l'approche « QTL », peut valider l'hypothèse selon laquelle le locus agit dans le processus d'adaptation (Rogers et Bernatchez 2005). Inversement, on peut tester la sélection de polymorphisme de gènes dont l'implication dans l'architecture du caractère adaptatif a été préalablement validée par génétique d'association.

### **I-2.2 Des avancées, mais encore de nombreuses questions**

Les approches précédentes ont connu un incroyable succès ces dernières années pour un grand nombre d'organismes et ont permis de dégager certaines généralités sur les bases génétiques de l'adaptation. Néanmoins un certain nombre de questions restent encore en suspens, en raison bien sûr des limitations inhérentes aux différentes approches, mais aussi de la complexité de la réponse mise en œuvre face à la sélection. Voici donc une synthèse des principales interrogations soulevées par ces d'études.

### **I-2.2.1 Nombre et effets des gènes impliqués dans l'adaptation**

La complexité génétique des caractères héritable suggère que la réponse à la sélection peut se faire selon plusieurs modes : l'adaptation d'un caractère peut être due *a priori* à l'évolution d'un faible nombre de gènes, voire d'un seul, contribuant majoritairement aux modifications de ce caractère ou de plusieurs gènes dont les effets peuvent être faibles séparément ou importants lorsqu'ils s'ajoutent.

Historiquement, l'estimation des effets des gènes impliqués dans la variation de caractères soumis à sélection a fait l'objet de développements de modèles théoriques depuis les années 30 avec le modèle géométrique de Fisher (Fisher (1930) cité dans Orr (2005)). Dans ce modèle qui vise à formaliser l'évolution phénotypique des caractères, Fisher considère l'évolution phénotypique d'un individu éloigné de l'optimum qui peut s'en approcher par des modifications adaptatives. Ce modèle suggère que la probabilité qu'une modification soit avantageuse est plus grande lorsque les effets sur le phénotype sont petits, tandis que les changements phénotypiques de forts effets ont de plus grandes chances d'être délétères. Fisher en a conclu qu'un caractère évolue par une succession de petites modifications avantageuses (infinitésimales). Constatant que ces attendus théoriques sont en conflit avec les observations expérimentales, qui tendent à montrer que des gènes à forts effets peuvent être impliqués dans l'adaptation, Orr (2005) suggère que les bases génétiques responsables de l'évolution d'un caractère dépendent de l'éloignement à l'optimum du phénotype considéré. Sans modifier les hypothèses du modèle de Fisher, mais en considérant qu'un organisme évolue par mutations, il montre notamment que plus un phénotype est éloigné de l'optimum, plus il est probable que des mutations de forts effets soient avantageuses et se fixent, celles d'effets faibles étant plus facilement perdues par dérive. En revanche, à l'approche du phénotype optimal, les mutations de faibles effets sont plus facilement fixées (Orr 1998). Ce modèle suggère donc que la sélection d'un caractère devrait en premier lieu à conduire à la fixation d'allèles à effets forts sur le phénotype puis, dans un second temps, à celles des allèles de plus faibles effets. On s'attendrait donc à ce qu'au début de l'adaptation quelques gènes d'effets forts évoluent rapidement, puis que de petites modifications dans un grand nombre gènes ajustent plus finement le caractère près de l'optimum.

D'un point de vue expérimental, les détections de QTL pour différentes espèces tendent à montrer que le nombre de QTL identifiés pour un caractère adaptatif donné est variable mais reste assez faible en moyenne, allant de 1 à 10 par caractère et par espèce, tandis que leurs effets sont souvent assez forts expliquant jusqu'à 50% de la variance phénotypique (Erickson

*et al.* 2004). Les études portant sur la détection de locus sous sélection divergente sur l'ensemble du génome vont dans le même sens avec seulement 1% à 5% de locus trouvés dans la plupart des cas. Ainsi, dans une étude analysant les bases génétiques de l'adaptation de la grenouille commune le long d'un gradient altitudinal, sur les 392 marqueurs AFLP testés (Bonin *et al.* 2006), seuls 8 présentent des signatures de sélection et sont trouvés comme potentiellement impliqués dans l'adaptation à l'altitude. Cependant ces résultats expérimentaux doivent être interprétés avec précaution puisque : (i) Les régions à effets forts détectées comme impliquées dans la variation d'un caractère adaptatif par les méthodes de détection de QTL peuvent contenir plusieurs gènes de faibles effets. (ii) La détection de QTL à partir d'un nombre d'individus trop faible (inférieur à 300 individus) peut entraîner une mauvaise estimation de la distribution des effets des QTL. En particulier, l'effet fort des QTL détectés tend à être surestimé tandis que la puissance de détection des QTL de faible effet est fortement réduite (Erickson *et al.* 2004). (iii) Des études analytiques (Latta 1998 ; McKay et Latta 2002) ou basées sur des simulations (Le Corre et Kremer 2003) ont montré que la sélection d'un caractère gouverné par un grand nombre de locus à effets additifs tend à générer des covariances entre locus. La sélection sur le caractère se traduit alors par de faibles changements de fréquences alléliques aux locus pris individuellement de sorte que les locus ne sont pas détectés comme sélectionnés par les méthodes classiques de détection de la sélection présentées plus haut.

Dans une étude combinant la recherche de régions du génome sous sélection avec une étude de détection de QTL pour identifier les gènes responsables de la différenciation entre les formes naines et normales d'un poisson d'eau douce, le grand corégone, Rogers et Bernatchez (2005) ont trouvés 8 QTL potentiellement sous sélection sur les 27 détectés. L'absence de sélection à certains locus participant à la variation du caractère suppose que tous les locus participant à l'architecture d'un caractère ne sont pas soumis à la sélection directionnelle par exemple du fait de forces contradictoires agissant à ces locus (pléiotropie antagoniste). Cependant, on ne peut exclure l'existence de limites à la détection des traces de sélection à certains locus par les méthodes précédentes (*cf.* point (iii), et paragraphe suivant).

### **I-2.2.2 « Standing genetic variation » ou nouvelles mutations ?**

Les modèles théoriques de l'adaptation présentés précédemment (Fisher 1930 ; Orr 2005) ou ceux développés pour prédire l'évolution de la diversité autour d'un gène sélectionné

(phénomène d'auto-stop génétique ou balayage sélectif déjà mentionné précédemment, Maynard Smith et Haigh, 1974) supposent que l'évolution d'un caractère joue sur l'apparition de nouvelles mutations avantageuses fixées par sélection. Cependant lorsqu'on considère une population colonisant un nouvel environnement ou soumise à de nouvelles pressions de sélection, il semble que la sélection à court terme joue préférentiellement sur la variabilité disponible dans la population de départ (appelée « standing genetic variation », Barrett et Schluter, 2008). Du fait de sa présence en plus grande fréquence initiale, la variabilité disponible dans la population de départ présente une probabilité de fixation plus importante que les nouvelles mutations présentes en faible fréquence et pouvant être perdues rapidement par dérive, d'autant plus vite que l'effet de l'allèle sur le phénotype est faible (Figure I-1). Ainsi l'adaptation à partir de la variation initialement présente dans la population peut conduire à la sélection d'allèles de plus faibles effets que la sélection à partir de nouvelles mutations. La variation disponible dans la population permet également une réponse à la sélection *a priori* plus rapide dans un nouvel environnement comparé à la mutation car la variabilité est directement disponible pour répondre aux nouvelles pressions de sélection.

Une étude simulant l'évolution d'un jeu de données en considérant à la fois la sélection de nouvelles mutations et de mutations initialement présentes montre que les méthodes de « genome scan » ne permettent généralement pas de détecter la sélection à partir de la variation initialement présente (Teshima *et al.* 2006). En effet les méthodes de détections de sélection type « genome scan » sont basées sur la détection de régions présentant des signatures de balayage sélectif, c'est-à-dire une réduction de la diversité accompagnée d'un renforcement du déséquilibre de liaison autour du locus sélectionné. Cette signature est celle attendue dans le cas d'une sélection forte conduisant à la fixation rapide d'une nouvelle mutation. La sélection d'une mutation initialement présente dans la population avant sélection conduit à des signatures beaucoup moins marquées (Przeworski *et al.* 2005) qu'il n'est alors pas possible de détecter par les approches classiques.

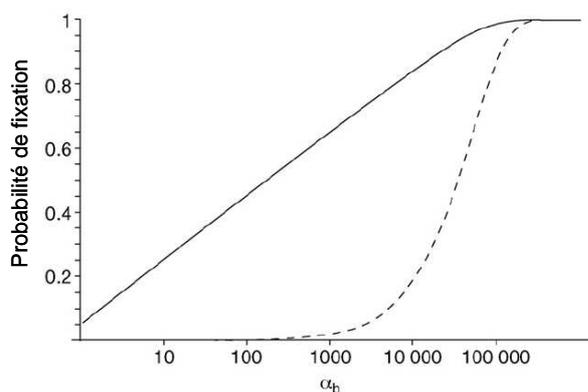


Figure I-1 : Probabilité de fixation d'une mutation avantageuse apparaissant lors d'une modification environnementale (courbe pointillée) et probabilité moyenne de fixation d'une mutation déjà présente lors d'une modification environnementale (courbe pleine) – neutre avant le changement et avantageuse après – en fonction de l'avantage sélectif ( $\alpha_b = 2Ne \cdot s_b$  où  $Ne$  est l'effectif efficace de la population et  $s_b$  son effet sur la valeur sélective) (source Barrett et Schluter, 2008).

### **I-2.2.3 Rôle de la pléiotropie**

Une des raisons pour lesquelles un gène impliqué dans le déterminisme d'un caractère adaptatif peut ne pas répondre à la sélection est qu'un même gène peut agir sur l'expression de plusieurs caractères à la fois, on parle alors de pléiotropie. La pléiotropie génère des corrélations génétiques entre caractères qui peuvent altérer ou contraindre la réponse à la sélection (dans le cas d'une pléiotropie dite « antagoniste »). Bien que ces corrélations soient prises en compte depuis longtemps dans les études de sélection phénotypique multi-caractères (Lande et Arnold 1983), leurs bases génétiques sont encore mal connues. Gardner et Latta (2007) soulignent par ailleurs que les deux principales causes possibles de corrélation génétique entre caractères, pléiotropie et déséquilibre de liaison, sont souvent difficiles à dissocier. En ce qui concerne la précocité de floraison, dont les bases génétiques sont maintenant bien connues chez *Arabidopsis thaliana*, Roux *et al.* (2006) listent l'ensemble des gènes impliqués dans le déterminisme du caractère, identifiés par génomique fonctionnelle, et qui représentent des cibles potentielles de la sélection pour une plus grande précocité. Parmi les 44 gènes identifiés, 24 ont des effets pléiotropes sur d'autres caractères et parmi ceux-ci, 14 présentant des effets pléiotropes sur la production de graines ont été identifiés comme ayant une faible probabilité de fixation en cas de sélection pour une plus grande précocité.

### **I-2.2.4 Rôle de l'épistasie**

Contrairement aux gènes à bases génétiques additives pour lesquels l'effet d'un allèle est le même quelque soit le fond génétique dans lequel il se trouve, l'effet des gènes montrant de l'épistasie varie en fonction des gènes avec lesquels ils sont associés. Les modèles théoriques de l'adaptation présentés précédemment ne considèrent que l'hypothèse simplificatrice de mutations à effets additifs sur les phénotypes. Mais l'épistasie doit *a priori* jouer un rôle important dans la réponse à la sélection de gènes dont les effets dépendent du fond génétique, lui-même susceptible d'évoluer du fait de la sélection (Hansen 2006). En raison de la difficulté à quantifier l'épistasie lors d'expérience de détection de QTL en particulier (Flint et Mott 2001), son effet sur la réponse à la sélection a été peu étudié expérimentalement (Demuth et Wade 2006) et il est pour l'heure difficile de conclure sur son implication dans l'architecture des caractères adaptatifs. Pourtant, les expérimentations effectuées

spécifiquement dans le but de détecter de l'épistasie semblent montrer que les interactions entre gènes jouent un rôle important dans l'architecture des caractères (Mackay 2001).

### **I-2.2.5 Type des mutations préférentiellement sélectionnées**

Une fois que les gènes soumis à sélection sont identifiés, on peut également s'intéresser plus précisément aux cibles des mutations impliquées dans l'adaptation. En particulier la sélection agit-elle préférentiellement sur des mutations affectant directement les régions codantes du gène touchant à sa fonction propre ou sur des mutations des régions régulatrices modifiant l'expression des gènes ? Les études comparatives tendent à montrer que les divergences protéiques observées ne sont pas suffisantes pour expliquer les différences entre phénotypes, ce qui suggère que dans de nombreux cas l'adaptation résulte de modifications dans l'expression des gènes plutôt que de modifications de leur fonction (Fay et Wittkopp 2008). Selon Wray (2007), deux raisons peuvent être invoquées pour expliquer l'importance des zones régulatrices. (i) Certains phénotypes peuvent être plus facilement obtenus par des mutations dans des régions régulatrices que dans des régions codantes. L'observation de l'évolution parallèle de la persistance de la lactase dans 2 populations humaines distinctes liée à des mutations indépendantes dans des régions régulatrices constitue un des exemples allant dans le sens de cet argument (Tishkoff *et al.* 2006) ; (ii) Les mutations dans les zones régulatrices peuvent réduire la pléiotropie, rendant la sélection plus efficace dans ces zones que dans les régions codantes. Cependant la validation de ces hypothèses nécessite de plus amples travaux expérimentaux, l'importance des mutations des zones régulatrices dans l'évolution phénotypique étant encore largement contestée (Hoekstra et Coyne 2007).

### **I-2.2.6 Conclusions**

Malgré des modèles théoriques développés et les nombreuses études analytiques visant à appréhender les bases génétiques de l'adaptation qui ont vu le jour avec l'émergence de la génomique, de nombreux aspects font encore débat, en raison notamment des limitations des techniques utilisées face à l'extraordinaire complexité de la relation phénotype-génotype-environnement. Etudier l'évolution d'un caractère adaptatif chez une espèce modèle, pour lesquelles des outils de génomique ont été développés, est alors une alternative puissante, complémentaires des précédentes approches, pour tenter de répondre aux questions que

soulève l'adaptation. L'utilisation de populations expérimentales, évoluant en temps réel, sous l'effet de la sélection naturelle et pour lesquelles certains paramètres démographiques peuvent être contrôlés, constitue un outil de choix pour étudier les mécanismes de l'adaptation. En effet, on peut ainsi disposer de nombreuses informations sur l'état de l'environnement, le contexte génétique, les valeurs phénotypiques et la taille d'une population, tout au long de son adaptation en milieu naturel. Ces informations sont justement celles qui font souvent défaut dans les méthodes précédentes et c'est cette approche qui est proposée dans le cadre de ma thèse.

## **I-3 Etude de l'adaptation locale dans des populations expérimentales de blé hexaploïde : exemple de la précocité de floraison**

### **I-3.1 Présentation des populations expérimentales de gestion dynamique**

#### **I-3.1.1 Principe des populations**

Les populations expérimentales considérées dans ce travail sont issues du programme de gestion dynamique des ressources génétiques du blé tendre (*Triticum aestivum* L.) mis en place en 1984 par l'INRA (Institut National de Recherches Agronomiques) et la DGER (Direction Générale de l'Enseignement et de la Recherche) du Ministère de l'Agriculture avec l'appui du BRG (Bureau des Ressources Génétiques). Ce programme a été initié avec un objectif double :

(i) Mettre en place un programme pilote de gestion des ressources génétiques complémentaire de la conservation dite « statique » des ressources dans des banques de graines : Le mode de conservation dynamique vise maintenir le potentiel adaptatif d'espèces d'intérêt agricole en conservant des populations génétiquement hétérogènes dans leurs agro-écosystèmes habituels où elles évoluent sous l'action des pressions évolutives (mutation, migration, sélection naturelle, dérive) voire également sous l'action de la sélection artificielle. Les programmes de conservation dynamique passent par la conservation à la ferme, par des expériences de sélection participative ou encore par la création de réseaux de gestion dynamique. Les

populations expérimentales étudiées pendant ma thèse ont été constituées à partir d'une même population initiale issue du croisement de plusieurs variétés de blé semée ensuite dans différents sites de culture en France. Trois pools initiaux ont été utilisés pour la mise en place du réseau de gestion dynamique : 2 populations PA et PB issues du croisement de 16 parents préférentiellement autogames (Thomas *et al.* 1991 ; Picard *et al.* 1984), et une population PS, issue d'inter-croisements entre 62 lignées d'origines diverses et rendue allogame par l'introduction d'un gène nucléaire de stérilité mâle récessif (Trottet, 1988). Toutes les populations ont évolué de façon indépendante et isolée dans les différents sites sans sélection humaine consciente et avec une démographie constante d'une année à l'autre. En l'absence de migration, sous l'effet de la sélection naturelle et de la dérive, la variabilité devrait diminuer à chaque site alors que la variabilité totale (à l'échelle du réseau) devrait être conservée du fait de la différenciation génétique des populations.

(ii) Etudier les mécanismes d'adaptation de populations d'une même composition initiale dans des environnements contrastés (Henry *et al.* 1991): Les populations expérimentales de GD constituent un outil puissant pour étudier la réponse mise en œuvre dans des populations soumises à des pressions de sélection différentes. L'absence de migration entre les populations permet d'étudier les seuls rôles de la dérive et de la sélection dans l'évolution des populations. Par ailleurs, l'utilisation d'une espèce annuelle implique que les générations ne se chevauchent pas, on peut donc ainsi étudier l'état de la population d'une génération à l'autre. Enfin la composition initiale des populations est connue, ce qui permet le suivi des fréquences alléliques aux gènes depuis le début de l'expérimentation jusqu'à une génération donnée.

### **I-3.1.2 Principaux résultats**

Les études précédentes ont mis en évidence une différenciation très rapide des populations par rapport à leur état initial. Deux grandes tendances d'évolution des caractères agromorphologiques ont été observées :

- Adaptation aux conditions de compétition : Une augmentation généralisée de la taille moyenne des plantes a été observée dans toutes les populations des pools PA et PB. Cette évolution unidirectionnelle est interprétée comme une réponse adaptative à des conditions de forte compétition pour la lumière.

- Adaptation aux milieux physique et biotique : David *et al.* (1992) ont montré que les populations du sud devenaient génétiquement plus précoces que les populations du nord dans les populations autogames PA : une corrélation positive a été mise en évidence entre la latitude du site de culture et la date d'épiaison dès la génération 8. Cette évolution est interprétée comme le résultat d'une adaptation des populations aux conditions climatiques. Par ailleurs, Lebouc'h *et al.* (1994) ont mis en évidence une différenciation nette pour la résistance à l'oïdium en fonction du site de multiplication résultant de l'adaptation des populations aux pressions pathogènes. Ceci se traduit par l'apparition de nouvelles combinaisons de gènes de résistance selon le site et un niveau variable de résistance adulte entre populations. Paillard (1999) a montré une augmentation de la fréquence de certains gènes de résistance aux rouilles en réponse aux pressions parasitaires. Son étude met également en évidence que la variabilité des gènes spécifiques de résistance à l'oïdium et aux rouilles est maintenue sur l'ensemble du réseau.

Encadré I-1 : Présentation du blé tendre.

❖ *Triticum aestivum* L., famille des Poacées

❖ Plante herbacée annuelle cultivée

❖ Espèce **autogame** au génome **allo-hexaploïde** ( $2n=42$ , génome AABBDD)

❖ Origine du blé tendre :

Originaire du croissant fertile (Irak actuel), le blé tendre a été domestiqué il y a 10000 ans

❖ Culture et débouché :

Le blé tendre est cultivé dans une très grande gamme de milieux. On distingue 2 types de blé :

- le **blé d'hiver** nécessitant un période de froid (vernalisation) pour initier la floraison. Ce type de blé est semé à l'automne dans les régions méditerranéennes et tempérées.
- le **blé de printemps** insensibles à la vernalisation. Ce type de blé est semé au printemps dans les pays à hiver rigoureux ou à l'automne dans les sites où les températures hivernales sont modérées.

De plus, le blé fleurit généralement en conditions de **jours longs** mais il existe des blés insensibles ce signal environnemental.

Le blé tendre **quatrième culture mondiale** avec une production de 626 Millions de tonnes dans le monde dont 24,8 millions de tonnes produits en France en 2005 (FAOSTAT [www.Fao.org](http://www.Fao.org)). Les débouchés sont l'alimentation humaine (pain, pâtisserie), l'alimentation animale et dans une moindre mesure l'industrie (papeterie, chimie, pharmacie)

### **I-3.3 Pourquoi s'intéresser à la précocité de floraison en particulier ?**

#### **I-3.3.1 La précocité de floraison, un caractère adaptatif majeur.**

La précocité de floraison est un caractère d'adaptation majeur chez les plantes annuelles telles que le blé et c'est d'ailleurs l'un des caractères, avec la résistance aux maladies, qui montre une évolution divergente dans les populations de GD.

Depuis le début de l'agriculture, ce caractère a été sélectionné artificiellement chez les espèces cultivées afin de permettre leur culture dans un nouvel environnement. L'exemple le mieux documenté est peut-être celui du maïs dont l'aire d'origine se situe au Mexique au climat de type tropical et qui est actuellement cultivé sur une grande partie du continent américain et en Europe sous des climats tempérés. Rebourg *et al.* (2003) ont montré que la culture dans des conditions tempérées a été rendue possible en sélectionnant des variétés insensibles à la photopériode, extrêmement précoces, capables d'accomplir leur cycle sur une période de temps réduite permettant une culture limitée à la période chaude de ces latitudes (semis en mai, récolte en octobre).

L'étude de caractère adaptatif, tel que la précocité de floraison, est par ailleurs intéressante pour appréhender les réponses adaptatives qui seront mises en œuvre face aux changements environnementaux futurs. Dans le passé, les plantes ont répondu aux modifications du climat par l'occupation progressive de nouvelles aires de distribution. Cependant du fait de la rapidité des changements climatiques contemporains attendus, et de la fragmentation des habitats qui réduit les possibilités de migration, il est probable que seules les espèces pouvant s'adapter en modifiant leur cycle de vie pourront se maintenir. Un certain nombre d'études expérimentales vont dans ce sens, notamment une étude récente portant sur les conséquences du réchauffement climatique à l'échelle planétaire en simulant des populations soumises à des températures plus élevées que celles observées dans leurs conditions naturelles. Dans cette étude, Franks *et al.* (2007) tendent à montrer que la réponse adaptative mise en œuvre chez une espèce annuelle, *Brassica rapa*, passe par un raccourcissement de la phase végétative et une précocité de floraison accrue des populations.

### **I-3.3.2 L'architecture du caractère de précocité chez le blé**

Les variétés cultivées de blé tendre, comme la plupart des céréales de milieux tempérés, peuvent être caractérisées en fonction de leur sensibilité au froid (vernalisation) et à la longueur du jour (photopériode) pour l'initiation de la floraison (*cf.* Encadré I-1). Très tôt, l'utilisation de lignées de substitution et de croisements contrôlés a permis de mettre en évidence l'existence de gènes majeurs impliqués dans la réponse à ces deux signaux environnementaux. Par la suite, trois voies de régulation de la précocité de floraison ont été identifiées (Worland 1996) : la voie de la vernalisation, la voie de la photopériode et la voie de la précocité intrinsèque. Si des gènes majeurs ont été trouvés comme étant impliqués dans les voies de la vernalisation (gène *VRN-1*) et de la photopériode (gène *PPD-1*), la voie de la précocité intrinsèque, qui détermine la date de floraison indépendamment des stimulus environnementaux, semble impliquer un plus grand nombre de gènes de plus faibles effets. Par la suite, les études de cartographies de QTL ont permis de localiser plus précisément le positionnement de ces gènes sur le génome du blé (Sourdille *et al.* 2000 ; Kuchel *et al.* 2006 ; Hanocq *et al.* 2007). Notons que du fait de la nature allo-hexaploïde du génome du blé, chaque gène est potentiellement présent en 3 copies homéologues distinctes ayant peu divergées. Cette architecture du génome complique l'étude individuelle des gènes et ce n'est que très récemment et en ce basant sur l'étude préliminaire des espèces diploïdes apparentées que les 2 gènes majeurs ont pu être isolés et clonés (Yan *et al.* 2003 ; Beales *et al.* 2007). Des zones polymorphes ont été trouvées dans les régions régulatrices ou dans les régions codantes de ces gènes, qui varient avec la précocité de floraison dans différentes collections (Yan *et al.* 2004 ; Sherman *et al.* 2004 ; Fu *et al.* 2005 ; Beales *et al.* 2007).

### **I-3.3.3 Déterminisme bien connu pour des espèces modèles tels le riz ou *Arabidopsis thaliana***

Comme en témoignent la distribution continue du caractère et les études de détection de QTL, l'architecture de la précocité chez le blé ne se résume pas à ces deux gènes majeurs mais implique également un grand nombre d'autres gènes en interaction et à effets plus modérés.

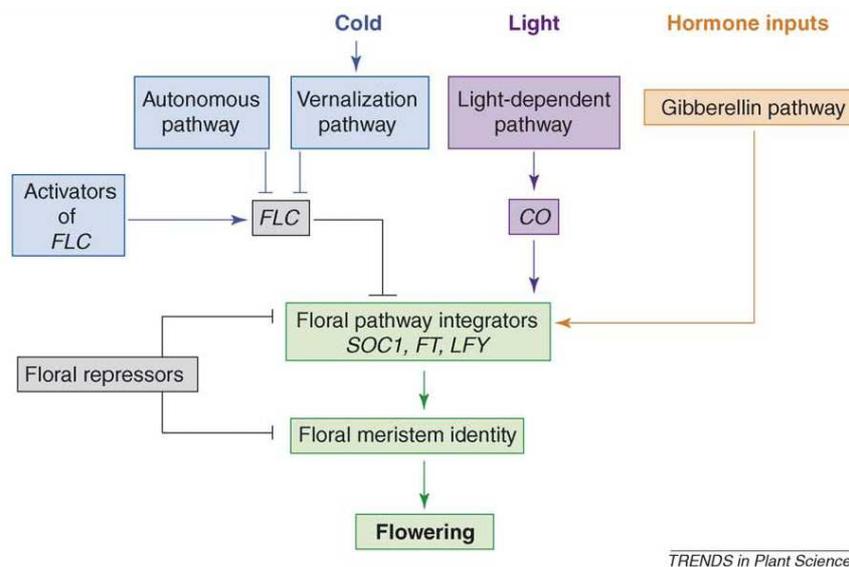


Figure I-2 : Représentation schématique des voies de signalisation impliquées dans la floraison de *Arabidopsis thaliana* et leur interactions (Source : Roux *et al.* 2006).

La recherche d'homologies de séquences avec des séquences de gènes connus chez les espèces modèles est une stratégie communément utilisée pour isoler des gènes candidats. Le riz, pour lequel un grand nombre de ressources génomiques est disponible, peut être particulièrement intéressant du fait de sa proximité phylogénétique avec le blé, mais ces deux espèces ont développé des stratégies de développement différentes (le riz fleurit dans des conditions de jours courts et ne nécessite pas de période de vernalisation) qui peuvent compromettre la comparaison (Cockram *et al.* 2007). En revanche, bien que plus éloignée phylogénétiquement, *Arabidopsis thaliana*, qui répond aux mêmes signaux environnementaux que le blé, constitue une bonne alternative pour identifier des gènes orthologues candidats. Le déterminisme de la floraison est bien connu chez cette espèce modèle, avec plus de 80 gènes identifiés (Mouradov *et al.* 2002; Putterill *et al.* 2004). Ce type d'approche a été initié au laboratoire de génétique végétale du Moulon avant le début de cette thèse et a permis l'identification de polymorphismes dans des gènes candidats orthologues de gènes d'*Arabidopsis thaliana* et de riz appartenant aux différentes voies de régulation (*FT*, *GI*, *CO*, *LD*). Cependant, outre le problème méthodologique posé par la polyploïdie pour distinguer les locus homéologues, la diversité nucléotidique du blé autogame (environ 1 SNP toutes les 300 paires de bases en moyenne) est réduite par rapport à celle d'espèces allogames telles que le maïs (environ 1 SNP toutes les 100 paires de bases en moyenne). Par ailleurs, la domestication et les cycles de sélection conduits pour créer de nouvelles variétés ont probablement amplifié ce phénomène, puisqu'une étude récente (Haudry *et al.* 2007) montrent que la diversité nucléotidique est réduite de 69% chez les blés cultivés comparée à la

diversité des blés sauvages. Ainsi, peu de polymorphismes candidats sont finalement disponibles à l'issue de l'approche gène candidat initiée.

L'association des polymorphismes de gènes avec la précocité a ensuite été testée en parallèle de ce travail de thèse, par génétique d'association dans une collection de 235 lignées d'origine géographique large et de statuts différents (variétés ou populations modernes ou anciennes) (Rousset *et al.* en préparation ; Bonnin *et al.* 2008). Finalement, 13 régions variables (SNP ou Insertion/Délétion), polymorphes dans les lignées parentales à l'origine des populations de GD, sont disponibles pour étudier l'évolution de gènes plus ou moins impliqués dans l'architecture de la précocité de floraison.

Tableau I-1 : Tableau de synthèse rassemblant des marqueurs de gènes candidats disponibles pour la thèse. ND : Non disponible.

		gènes	Publication associée	génom	taille du fragment de gène isolé (kb)	polymorphisme pour les 16 parents de GD	Association avec la précocité par génétique d'association
	Voie de la photopériode chez <i>A. thaliana</i>	<i>Constans (CO)</i> (CH5 bras long)	-	B	0.3	1 SNP (intron)	-
		<i>Gigantea (G1)</i> (CH3 bras court)	-	B	1.3	1 SNP (exon)	-
	Voie autonome chez <i>A. thaliana</i>	<i>Luminidependens (LD)</i> (CH3 bras long)	-	B	2	2 SNP espacés de 300pb (intron)	x
	gène intégrateur vernalisation/ photopériode chez <i>A. thaliana</i>	<i>Flowering locus T (FT)</i> (CH7 bras court)	Bonnin <i>et al.</i> (2008)	A	1	1 motif microsatellite et 1 SNP espacés de 200 pb (intron)	xx
D				1	1 indel (exon)	xx	
	Voie de la vernalisation	<i>VRN-1</i> (CH5 bras long)	Yan <i>et al.</i> (2004) Sherman <i>et al.</i> (2004)	A	12	Polymorphismes (région promotrice et exon 7 espacés de 9kb)	xxx
				B	12	délétion de 4 Kb (intron)	xxx
			Fu <i>et al.</i> (2005)	D	12	délétion de 4 Kb (intron)	xxx
	Voie de la photopériode	<i>Ppd-1</i> (CH2 bras court)	Beales <i>et al.</i> 2007	D	≈ 10	délétion de 2kb	ND
						délétion de 16pb (exon)	ND

## I-4 Objectifs et plan de la thèse

L'objectif de cette thèse est de comprendre quels ont été les mécanismes génétiques et sélectifs mis en jeu dans l'adaptation climatique de 3 populations expérimentales de blé ayant évolué pendant 12 générations dans des environnements contrastés. Du fait des ressources génomiques disponibles pour ce caractère, cette étude porte principalement sur l'évolution de la précocité bien que l'évolution d'autres caractères liés à la valeur sélective sera prise en considération pour tenir compte des corrélations possibles qui peuvent contraindre la réponse à la sélection.

Dans un premier temps, j'utiliserai des marqueurs microsatellites pour mettre en évidence le rôle de la dérive génétique et du système de reproduction sur l'évolution de la diversité neutre des populations indépendamment du caractère adaptatif étudié (Chapitre III). Puis je testerai différentes méthodes de détection de la sélection sur le caractère de précocité de floraison en me référant soit à l'évolution neutre attendue, soit à l'évolution de caractères liés à la valeur sélective tels que le nombre et le poids de descendant (Chapitre IV). Enfin, je mettrai en relation les variations de polymorphismes de gènes candidats avec l'évolution de la précocité (Chapitre V). Au fur et à mesure de l'étude, chacun de ces 3 niveaux de variations sont mis en relation pour qu'émerge une vision globale et intégrée des forces gouvernant l'évolution des populations (Figure I-3).

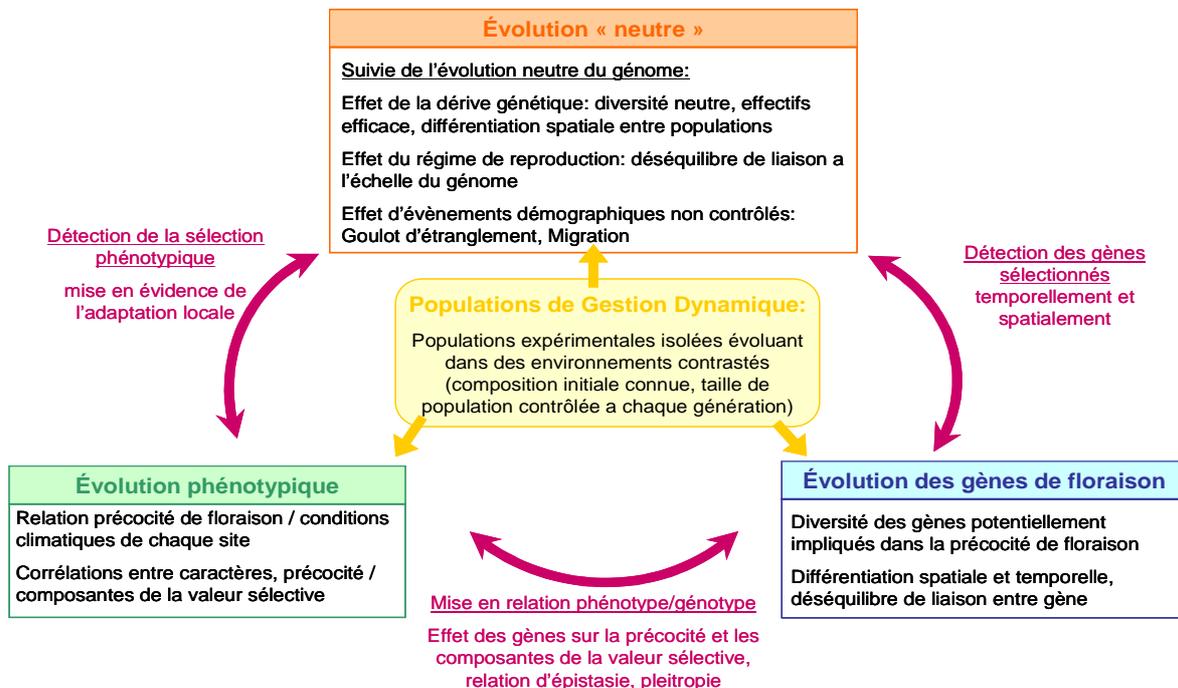


Figure I-3 : Représentation schématique des relations entre les différents niveaux de diversité et leur évolution mis en avant dans cette thèse.

## *Chapitre 2*

### *MATERIEL ET METHODES*

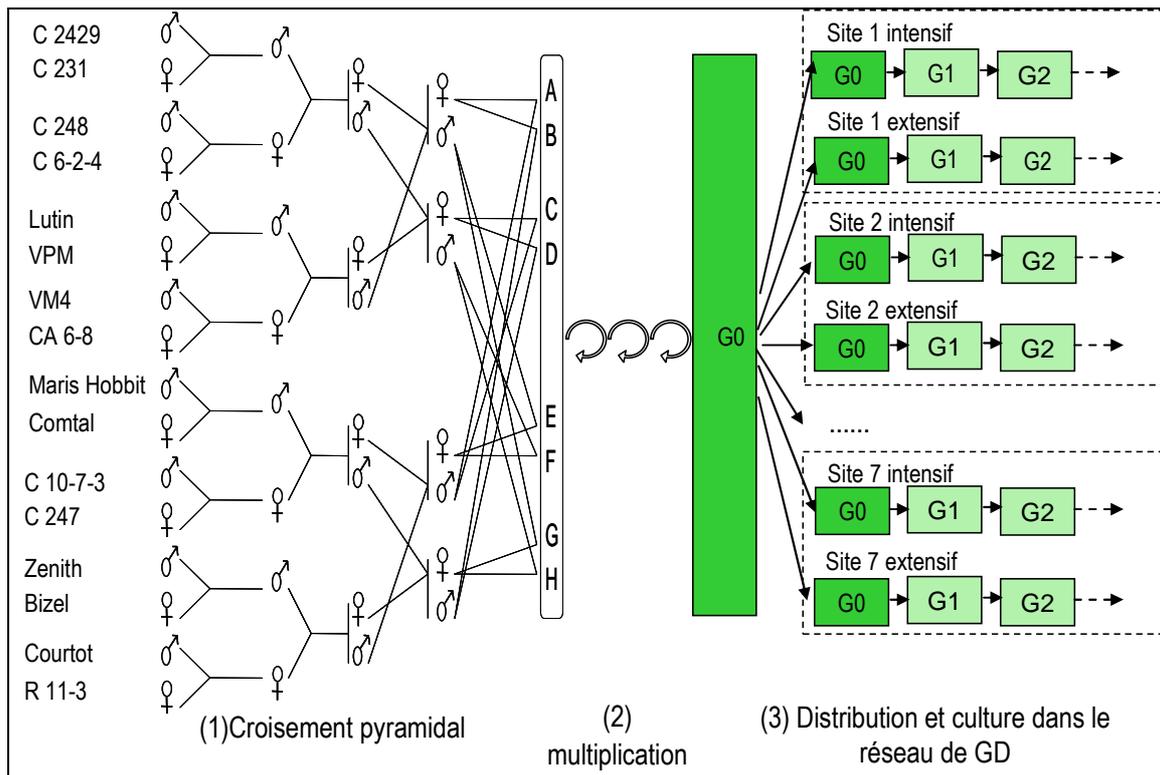


Figure II-1 : Les différentes étapes de la constitution des populations de GD : (1) 4 générations de croisement pyramidale des 16 lignées parentales; (2) 3 générations de multiplication en fécondation libre des semences ; (3) distribution de la population G0 dans les différents sites et culture en isolement.



Figure II-2 : Les 7 sites constituant le réseau français de gestion dynamique de ressources génétiques du blé tendre. Les sites indiqués en gras et soulignés sont ceux considérés dans cette étude.

## **Chapitre II : Matériel et méthodes**

Ce chapitre vise à présenter le matériel utilisé dans ce travail ainsi que l'acquisition des données brutes de caractérisation phénotypique et génotypique. Les méthodes statistiques utilisées pour l'analyse des données seront présentées par la suite au fur et à mesure dans la partie « Méthodes » de chaque chapitre.

### **II-1. Matériel végétal**

#### **II-1.1 Conduite des populations de gestion dynamique**

La population de gestion dynamique (GD) étudiée ici, nommée « PA », est issue du croisement de 16 lignées autogames de blé tendre d'hiver choisies pour leurs bonnes valeurs agronomiques (*i.e.* productivité élevée, bon rendement global) et la complémentarité de leurs caractères agronomiques (notamment résistances aux maladies, Thomas *et al.* 1991). Des croisements artificiels ont été effectués pendant 4 générations de façon pyramidale (croisements 2 à 2 successifs, Figure II-1). Ces générations d'inter-croisements ont été suivies par 3 générations de multiplication libre afin d'obtenir une population à base génétique large appelée G0 (ou PA0) et qui constitue la population initiale du programme de gestion dynamique. En 1984, des graines de la population initiale G0 ont été semées dans les sept sites du réseau de GD constitués de stations INRA et de lycées agricoles répartis sur l'ensemble du territoire français (Figure II-2). Puis les populations ont été cultivées de génération en génération sans sélection humaine consciente et sans migration entre populations. Dans chaque site, 2 populations ont été conduites sous 2 conditions de culture différentes : une condition dite « intensive » qui correspond aux pratiques conventionnelles de culture du blé tendre et une condition dite « extensive » avec un tiers de fertilisation azoté par rapport à la méthode conventionnelle et aucun traitement fongicide. Chaque année, l'ensemble de la parcelle a été récolté en mélange à maturité, un échantillon de 5000 à 10000 graines a été semé à l'automne pour constituer la génération suivante dans une parcelle de 100m<sup>2</sup> isolée d'autres cultures de blé et 500 à 1000 grammes de graines ont été conservés en chambre froide

à 4°C. Le stock ainsi conservé constitue une image de la composition de la population à une génération donnée et ce sont les semences issues de ces stocks qui sont utilisées pour caractériser la précocité des populations dans cette étude.

## **II-1.2 Echantillonnage**

Dans le cadre de ce travail, nous étudions l'évolution de la précocité de floraison et de gènes impliqués dans la variation de ce caractère dans (i) les 16 lignées parentales à l'origine des populations de GD, (ii) 36 génotypes de la population initiale G0 et (iii) 540 génotypes issus de trois populations évoluant en condition extensive dans des sites aux conditions climatiques contrastées (Vervins, noté VV, 49.5°N, 3.5°E ; Le Moulon, noté LM, 48.4°N, 1.4°W et Toulouse, noté TO, 43.4°N, 1.3°E, Figure II-2) et échantillonnés (60 génotypes par génération et par population) à des intervalles de temps réguliers : en génération 2 (G2, cultivée en 1986), en génération 7 (G7, 1991) et en génération 12 (G12, 1996). Au total, 592 génotypes ont donc été étudiés dans ce travail.

## **II-1.3 Constitution du matériel d'étude**

Les graines de la population initiale G0 utilisées ici sont issues d'au moins 3 générations d'autofécondation du stock initial. Ces semences ayant perdu au cours du temps leur pouvoir germinatif, seul un faible échantillonnage a pu être obtenu. Pour les 540 génotypes des trois sites de population de GD, les graines utilisées pour cette étude proviennent du stock de graines préservées au froid à chaque génération. Les 592 graines de l'expérimentation ont été semées au cours de l'automne 2003 sur le site du Moulon à faible densité. Pour chaque plante, correspondant à un génotype, un morceau de feuille a été prélevé pour une extraction d'ADN et la descendance issue d'autofécondation a été utilisée pour la caractérisation phénotypique des populations en conditions contrôlées (Figure II-3).

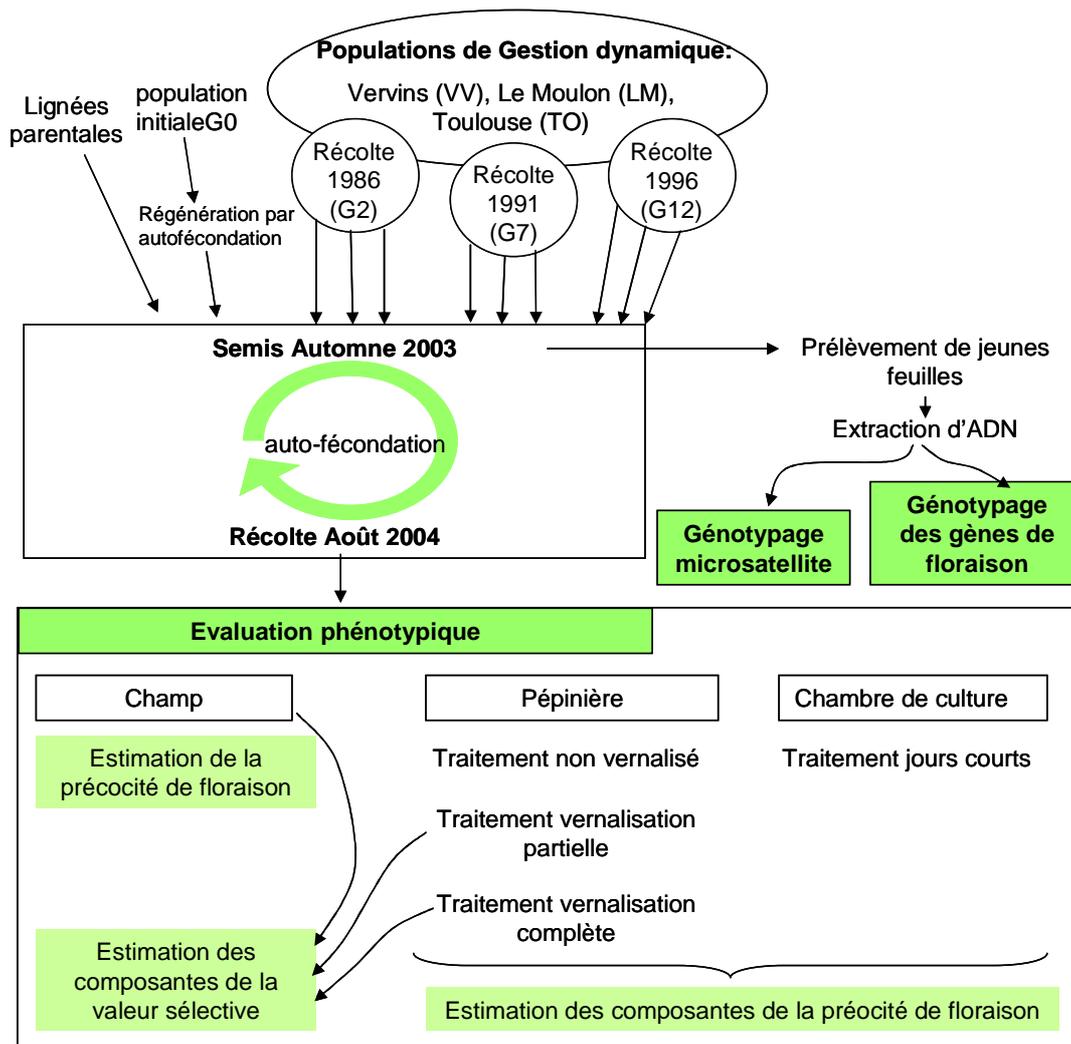


Figure II-3 : Constitution du matériel d'étude à partir des populations de gestion dynamique.

## II-2. Dispositif expérimental pour la caractérisation phénotypique des populations

### II-2.1 Estimation de la précocité de floraison au champ et sous différentes conditions de vernalisation en pépinière

#### - Dispositif au champ :

La précocité de floraison a été évaluée pour les 592 génotypes sur le site du Moulon dans un dispositif au champ où les plantes ont poussé dans des conditions « normales » de semis des populations de gestion dynamique (Figure II-4). Vingt graines de chaque génotype ont été semées sur une ligne de 1,20m avec 2 réplifications en randomisation totale début novembre

2004 et une seule réplique début novembre 2005. Pour chaque génotype, la date d'épiaison a été notée lorsque la moitié des épis de la ligne étaient à moitié sortis de leur gaine.

**- Dispositif en pépinière :**

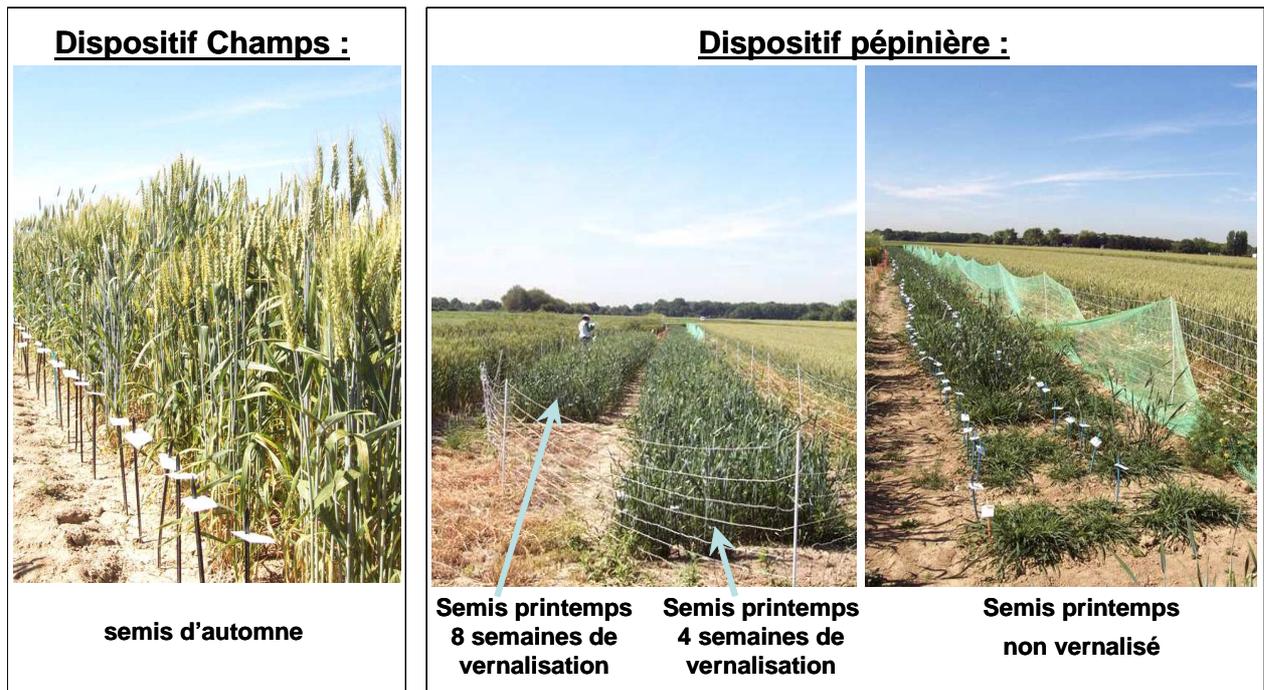
La précocité de floraison a également été évaluée sur les 592 génotypes qui ont subi différents traitements nous permettant de distinguer la contribution des différentes composantes influençant la floraison que sont les besoins en vernalisation, la sensibilité à la longueur du jour et la précocité intrinsèque :

- Traitement non vernalisé (NV) : six graines par génotype ont été semées dans des parcelles de 72m x 1,40m avec 2 génotypes par ligne de 1,40m dans la pépinière sur le site du Moulon durant le mois d'avril 2005. Les plantes se sont ainsi développées au printemps et en été sous des conditions de jours longs et n'ont pas subi de température vernalisante (*i.e.* inférieure à 10°C, Figure II-4).

- Traitement « vernalisation partielle » (4S) : des graines de chaque génotype ont été semées en terrine en mars 2005 et les plantules ont été mises en chambre de vernalisation (4°C, éclairage 12 heures) au stade « une feuille » pendant 4 semaines. Trois plantules par génotype ont ensuite été repiquées dans des parcelles de 52m x 1,40m avec 4 génotypes par ligne de 1,40m dans la pépinière sur le site du Moulon durant le mois d'avril 2005 (Figure II-4).

- Traitement « vernalisation complète » (8S) : le procédé est le même que précédemment : les graines ont été semées en février 2005 puis les plantules ont été mises à vernaliser pendant 8 semaines et repiquées dans la pépinière sur le site du Moulon au mois d'avril 2005 (Figure II-4).

Pour chacun des traitements, les génotypes ont été semés de façon complètement randomisée. La date de floraison a été notée sur chaque plante lorsque la moitié de l'épi le plus précoce était sortie de la gaine. L'évaluation des génotypes parentaux a été répétée en avril 2007 pour chacun des différents traitements.



*Photos B. Rhoné du 20/06/05*

Figure II-4 : Dispositifs au champ et en pépinière au Moulon pour estimer la précocité ainsi que les composantes de la précocité et de la valeur sélective.



Figure II-5 : Dispositif en chambre de culture en conditions de jours courts à Mons.

### **- Dispositif en chambres de culture :**

Afin d'affiner l'estimation des besoins en photopériode, une expérimentation en serre photopériodique à la station INRA de Mons-en-Chaussée (80, France) a été entreprise. En raison d'un manque de place dans les chambres de culture, l'expérimentation a été réalisée en deux parties : les génotypes parentaux ainsi qu'une partie des génotypes de la population G0 ont été évalués une première année en 2005 et le reste de la population a été évalué l'année suivante en répétant l'évaluation de 2 génotypes parentaux pour pouvoir comparer les données issues de l'expérimentation 2006 avec les données d'expérimentation de l'année 2005. Des graines de chaque génotype ont été semées en terrine en novembre et mises en chambre de vernalisation à l'INRA du Moulon (4°C, éclairage 12 heures) au stade « une feuille » pendant 8 semaines (vernalisation complète). Les plantules vernalisées ont été transférées à l'INRA de Mons en janvier et repiquées dans des pots de 2 litres, à raison de 2 plantes par génotype et 2 génotypes par pot. Afin de limiter la compétition entre génotypes et faciliter les notations, les génotypes repiqués ensemble dans un même pot ont été choisis de même classe de précocité (précoce, intermédiaire, tardif) en se fondant sur la précocité au champs et randomisés au sein de chaque classe. Deux plantes de chaque génotype ont été entreposées dans une chapelle de chambre de culture pour y subir un traitement de « jours courts » (JC) : les plantules ont subi un éclairage de 9h par jour, la nuit étant créée artificiellement par les écrans noirs de la serre (Figure II-5).

L'ensemble des notations a été effectué par le personnel de la station de Mons. La date de floraison a été notée plante à plante lorsque la moitié de l'épi le plus précoce était sortie de la gaine.

Les dates de floraison estimées dans les 3 dispositifs décrits ci-dessus ont été transformées en sommes de degrés-jours ( $^{\circ}\text{J}$ ) pour standardiser les notations effectuées dans les différentes conditions. Les mesures de précocité dans les dispositifs en pépinières et en chambres de culture ont été effectuées plante à plante pour les 6, 3 ou 2 plantes évaluées, la moyenne arithmétique par génotype a ensuite été calculée. Dans le dispositif en pépinière, l'expérimentation a été arrêtée le 1<sup>er</sup> août 2005 alors que certains génotypes n'avaient pas épié dans les traitements NV et 4S. Le nombre de degrés-jours cumulés à cette date est de 1892 $^{\circ}\text{J}$  et 1752 $^{\circ}\text{J}$  dans les traitements NV et 4S respectivement. Ainsi, pour les besoins des analyses statistiques, une valeur seuil d'épiaison de 2000 $^{\circ}\text{J}$ , a été attribuée aux génotypes non épiés à la fin de l'expérimentation (*cf.* Rhoné *et al.* 2008, chapitre 5). De même dans le dispositif en chambre de culture, l'expérimentation a été arrêtée le 4 juillet 2006 (2787 $^{\circ}\text{J}$  cumulés pour le

traitement jours courts). Près de la moitié des génotypes n'avaient pas épié à cette date dans les conditions de jours courts, une valeur seuil de 3000°J leur a été attribuée.

## II-2.2 Estimation des composantes de la précocité à partir des différents traitements

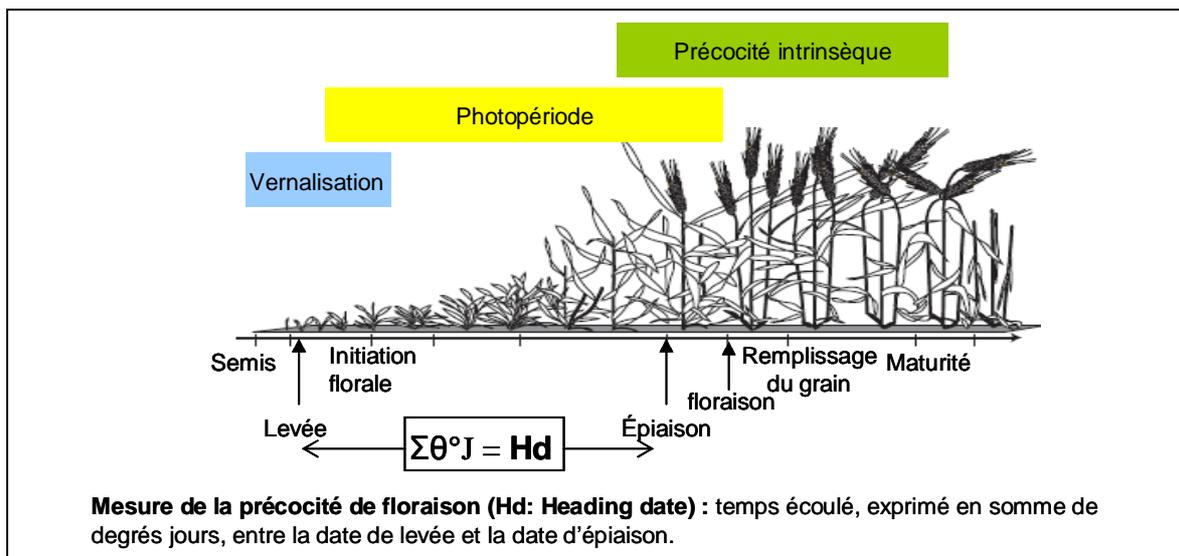


Figure II-6 : Décomposition des composantes de la précocité de floraison chez le blé tendre (source : Reynolds *et al.* 2001).

L'estimation des composantes de la précocité de floraison a été réalisée à partir des mesures de précocité effectuées dans les différents traitements (Figure II-6 ; Encadré II-1) :

- Besoins en vernalisation : le blé tendre a besoin d'être exposé à de faibles températures (environ inférieures à 10°C) pour initier la floraison. En l'absence de période de froid, la floraison est très fortement retardée pour les blés de type « hiver ». Les besoins en vernalisation ont été estimés par la mesure de précocité effectuée sur les plantes partiellement vernalisées pour lesquelles les besoins en photopériode sont satisfaits (dispositif 4S en pépinière sous jours longs). Le caractère « hiver » ou « printemps » des différents génotypes a été déterminé en se basant sur l'épiaison des plantes de chaque génotype en l'absence de vernalisation (Voir aussi Rhoné *et al.* 2008, chapitre 6).

- Besoins en photopériode : le blé tendre est une plante fleurissant sous jours longs (longueur du jour supérieure à 12 h). Les besoins en photopériode ont été estimés à partir de la mesure d'épiaison des plantes satisfaites pour les besoins en vernalisation mais poussant dans des conditions de courte longueur de jour (dispositif JC en chambre de culture).

- Précocité intrinsèque : une fois les besoins en photopériode et en vernalisation satisfaits, les variations de date de floraison observées entre génotypes sont attribuées à la précocité intrinsèque. La précocité intrinsèque a été estimée dans des conditions de vernalisation et de photopériode satisfaites (dispositif 8S en pépinière).

<b>Encadré II-1 : Méthode d'estimation des composantes de la précocité de floraison.</b>
<p>➤ <b>Détermination du type hiver/printemps:</b> Semis en pépinière au printemps (besoins en jours long satisfaits), traitement non vernalisé (NV): <math>\Sigma\theta^{\circ}\text{J} = \text{HdNV}</math></p> <div style="background-color: #ADD8E6; padding: 5px; text-align: center;"> <math>\text{HdNV} &lt; 1500^{\circ}\text{J} \rightarrow</math> type « printemps » (P)  <math>&gt; 1500^{\circ}\text{J} \rightarrow</math> type « hiver » (H)         </div>
<p>➤ <b>Détermination des besoins en vernalisation (BV):</b> Semis en pépinière au printemps (besoins en jours long satisfaits), traitement vernalisation partielle (4S): <math>\Sigma\theta^{\circ}\text{J} = \text{Hd4S}</math></p> <div style="background-color: #ADD8E6; padding: 5px; text-align: center;"> <math>\text{BV} = \text{Hd4S}</math> </div>
<p>➤ <b>Détermination des besoins en photopériode (BP):</b> Repiquage de plantes vernalisées en serres photopériodiques, traitement sous jours courts (JC) : <math>\Sigma\theta^{\circ}\text{J} = \text{HdJC}</math></p> <div style="background-color: #FFFF00; padding: 5px; text-align: center;"> <math>\text{BP} = \text{HdJC}</math> </div>
<p>➤ <b>Détermination de la précocité intrinsèque PI:</b> Semis en pépinière au printemps (besoins en jours longs et en vernalisation satisfaits) (8S): <math>\Sigma\theta^{\circ}\text{J} = \text{Hd8S}</math></p> <div style="background-color: #9ACD32; padding: 5px; text-align: center;"> <math>\text{PI} = \text{Hd8S}</math> </div>

## II-2.3 Estimation des composantes de la valeur sélective

Parmi les plantes évaluées en 2005 pour la précocité au champ, les mesures des composantes de la valeur sélective ont été effectuées sur 3 plantes par génotype, choisies au hasard dans le rang de 20 plantes de la première répétition. Ces mesures ont également été effectuées sur les 3 plantes des traitements 4S et 8S de la pépinière. Les caractères mesurés sont la hauteur de la plante entière à maturité, le nombre d'épis, le nombre de grains par plante et le poids de grains par plante.

## **II-3. Marquage microsatellite des populations de GD**

Les protocoles de biologie moléculaire utilisés dans cette thèse sont donnés en Annexe I-A.

### **II-3.1 Extraction**

Pour chacun des 592 géotypes utilisés dans cette étude, les extractions ont été effectuées à partir de 200mg de jeune feuille en suivant un protocole couramment utilisé dans le laboratoire adapté du protocole du kit d'extraction DNeasy 96 Plant de Qiagen (QIAGEN, Basel, Suisse). Les extractions ont été effectuées par l'équipe Biologie Moléculaire et Bioinformatique (BMBI) de l'INRA du Moulon et ont été utilisées pour l'ensemble des expérimentations de marquage moléculaire présentées par la suite.

### **II-3.2 Choix des marqueurs**

Pour étudier l'évolution de la diversité neutre dans les populations de GD, nous avons choisi d'utiliser des marqueurs de type « microsatellite ». Ce type de marqueurs présente l'avantage d'être abondant dans le génome, co-dominant, hautement polymorphe et impliquant des techniques de génotypage peu coûteuses et automatisables. De plus, de nombreux marqueurs microsatellites ont été développés et cartographiés chez le blé hexaploïde ces dernières années (Röder *et al.* 1998, Guyomarc'h *et al.* 2002, Somers *et al.* 2004, Song *et al.* 2005). D'autres types de marqueurs moins mutables, tels que les AFLP (Amplified Fragment Length Polymorphism) sont désormais très utilisés en génétique des populations, mais ces marqueurs sont dominants (détection des hétérozygotes impossible, données manquantes et allèles récessifs confondus). De nombreuses méthodes et outils statistiques ont récemment été développés permettant leur utilisation pour l'estimation des principaux paramètres de génétique des populations (Bonin *et al.* 2007), mais leur utilisation chez le blé reste délicate en raison de la nature hexaploïde du génome du blé.

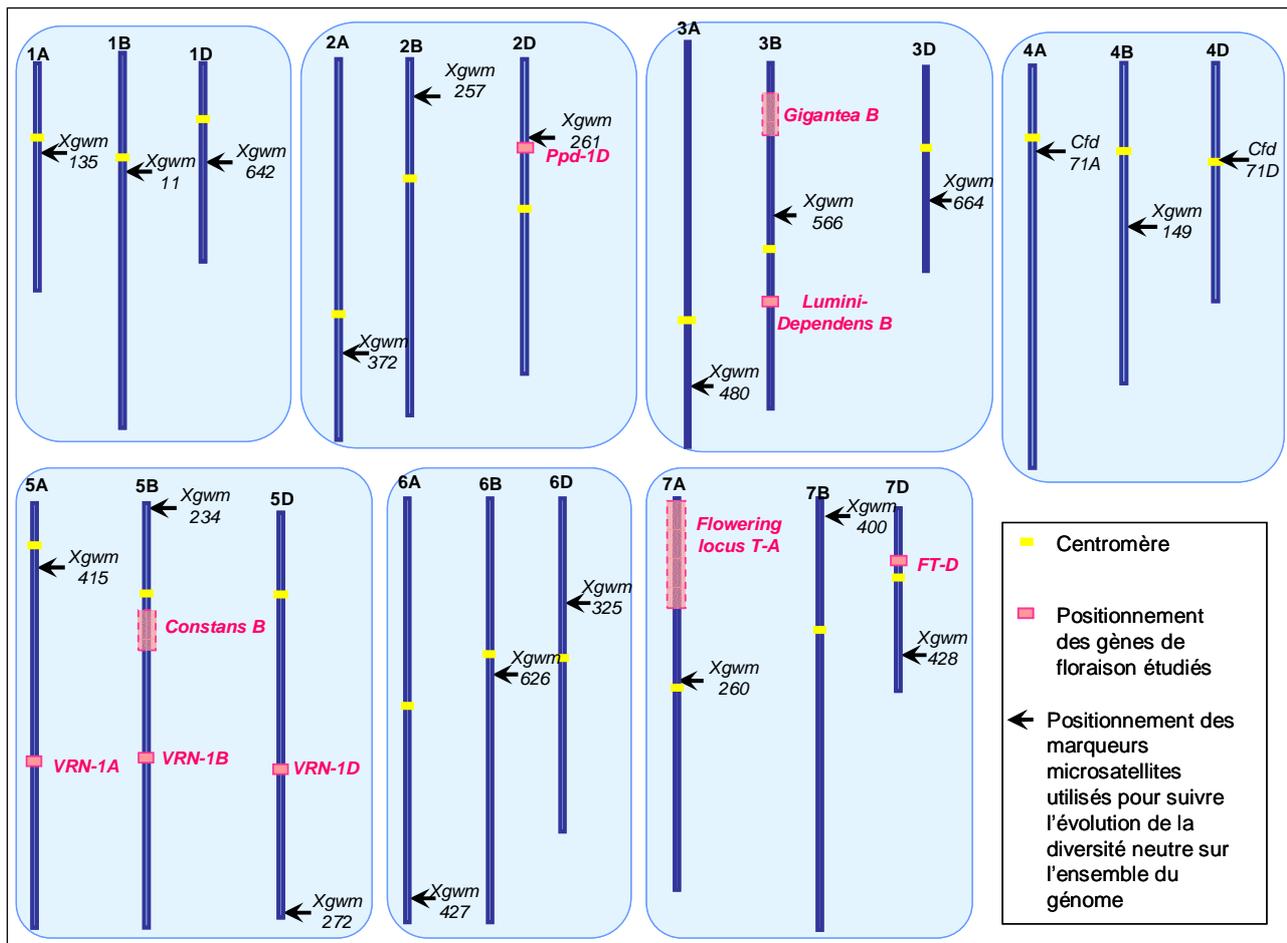
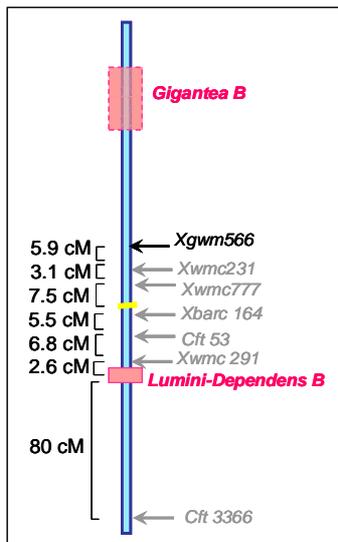


Figure II-7 : Positionnement des 20 marqueurs sur les 21 chromosomes du blé (d'après Somers *et al.* 2004, site internet graingenes (<http://wheat.pw.usda.gov/GG2/index.shtml>) et communication personnelle de Pierre Sourdille).

L'évolution neutre de l'ensemble du génome a été suivie à l'aide de 20 marqueurs microsatellites amplifiant 21 locus répartis sur l'ensemble du génome (1 locus par chromosome, Figure II-7). Les marqueurs choisis, développés par Röder *et al.* (1998, locus *Xgwm*) et Guyomarc'h *et al.* (2002, locus *Cfd71*), sont polymorphes chez les 16 lignées parentales de PA. La plupart de ces locus (15 sur les 20 marqueurs génotypés) ont été utilisés dans la thèse d'Anne-Laure Raquin (2005) portant sur la population de PA cultivée au Moulon.

Un suivi plus fin de l'évolution de la diversité neutre à l'échelle d'un chromosome a été entrepris. Sept marqueurs ont été choisis le long du chromosome 3B (dont celui utilisé pour étudier la diversité sur l'ensemble du génome, Figure II-8) selon des critères de polymorphismes dans la population parentale et de spécificité de l'amplification. En effet, de

part la nature hexaploïde du génome du blé, certaines amplifications non spécifiques du locus



étudié peuvent avoir lieu dans les régions homologues des autres chromosomes, brouillant ainsi le signal lors de l'électrophorèse. Les locus présentant des allèles nuls parmi les 16 lignées parentales ont également été écartés. Parmi les marqueurs retenus, 2 sont des marqueurs développés par Röder *et al.* (1998, *Xgwm*), 1 provient de Songs *et al.* (2002, *Xbarc*), 2 sont des marqueurs (*Cfd*) issus de d'extrémités de BAC développés à la station INRA de Clermont-Ferrand (63, France ; Communication P. Sourdille) et 3 proviennent de la carte consensus haute densité développée par Somers *et al.* (2004).

Figure II-8 : Positionnement des 7 marqueurs sur le chromosome 3B (d'après une communication personnelle de Pierre Sourdille).

### II-3.3 Génotypage et acquisition des résultats

Les amplifications par PCR (Polymerase Chain Reaction) ont été réalisées selon un protocole proche de celui proposé par les auteurs précédemment cités. Les produits amplifiés ont ensuite été séparés par électrophorèse sur gel d'acrylamide dans un séquenceur capillaire semi-automatique ABI 3100. L'amplification par PCR et la migration sur gel ont été réalisées à la plateforme INRA de génotypage à Clermont-Ferrand par le personnel de la plateforme.

J'ai ensuite analysé les résultats et réalisé les lectures à l'aide du logiciel d'analyse GeneMapper (v 3.7, Applied Biosystems).

## II-4 Génotypage des gènes candidats

### II-4.1 Révélation du polymorphisme de *VRN-1*

Quatre sites polymorphes chez les 16 lignées parentales PA des populations réparties sur les trois copies du gène *VRN-1* ont été génotypés pour ce travail :

- un polymorphisme de duplication, insertion et délétion dans le promoteur de la copie du génome A du gène *VRN-1A* (noté *VRN-1A<sub>pr</sub>* par la suite), révélé par Yan *et al.* (2004),
- une substitution dans l'exon 7 de *VRN-1A* (*VRN-1A<sub>ex7</sub>*), révélé par Sherman *et al.* (2004),
- une délétion de 4kb dans le premier intron de *VRN-1B* (*VRN-1B<sub>int1</sub>*) et dans le premier intron de *VRN-1D* (*VRN-1D<sub>int1</sub>*), toutes deux révélées par Fu *et al.* (2005).

Pour tous les sites polymorphes de *VRN-1*, des marqueurs co-dominants ont été développés (Yan *et al.* 2004, Sherman *et al.* 2004, Fu *et al.* 2005). Les conditions PCR ont été adaptées à partir des protocoles proposés par les auteurs et sont données en annexe. Pour détecter les allèles au locus *VRN-1A<sub>pr</sub>*, le génotypage a été effectué par migration sur gel d'acrylamide. L'amorce « forward » a été rallongée avec une queue M13 et les amplifications PCR ont été effectuées en présence de l'amorce M13 complémentaire marquée par fluorescence suivant le protocole proposé par Boutin-Ganache *et al.* (2001). Les produits d'amplification ont été séparés sur gel dénaturé de poly-acrylamide à 6.5% dans un séquenceur à plaques LI-COR (LI-COR Biosciences, Lincoln, USA). Les variations au locus *VRN-1A<sub>ex7</sub>* après digestion du produit de PCR (marqueur de type CAPS, Cleavage Amplification Polymorphic Sequence) et aux locus *VRN-1B<sub>int1</sub>* et *VRN-1D<sub>int1</sub>* (présence ou absence de délétions) ont été révélées par migration sur gel d'agarose à 2% et 0.8% respectivement et visualisées sous éclairage UV.

## II-4.2 Révélation du polymorphisme de *Ppd-1*

Deux sites polymorphes chez les 16 lignées parentales PA des populations tous deux localisés sur la copie du génome D du gène *Ppd-1* (Beales *et al.* 2007) ont été génotypés pour ce travail :

- une délétion localisée 2 kb en amont de la séquence codante (notée *Ppd-1D<sub>pr</sub>* par la suite)
- une délétion de 16 pb localisée dans l'exon 8 (*Ppd-1D<sub>ex8</sub>*) révélée par Beales *et al.* (2007).

Des marqueurs co-dominants permettant de révéler ces polymorphismes sur l'ensemble des populations étudiées ont été développés par les laboratoires de la société Biogemma à Riom (63, France). Les conditions PCR utilisées sont données en annexe. Les variations de *Ppd-1D<sub>pr</sub>* ont été révélées sur gel d'agarose 2%, tandis que pour *Ppd-1D<sub>ex8</sub>*, le polymorphisme a

été révélé par migration des amplifications PCR sur gel d'acrylamide dans un séquenceur capillaire automatisé ABI 3100. Le génotypage de ces deux régions polymorphes a été effectué dans les laboratoires de la société Biogemma avec l'aide du personnel technique du laboratoire.

### II-4.3 Génotypage de *FT*

Trois sites polymorphes chez les 16 lignées parentales PA des populations (Bonnin *et al.* 2008) répartis sur 2 des trois copies du gène *FT* ont été génotypés dans cette étude :

- un SNP (Single Nucleotide Polymorphism ; C/T) localisé dans le premier intron de *FT-A* (*FT-A<sub>int1</sub>*),
- un motif microsatellite variable ((TG)<sub>8ou9</sub>-(CG)<sub>3ou4</sub>) localisé dans le second intron de *FT-A* (*FT-A<sub>int2a</sub>* noté FTA (TG)<sub>8or9</sub> dans Bonnin *et al.* 2008, et *FT-A<sub>int2b</sub>*), à 130 pb du site polymorphe précédent,
- une insertion-délétion d'une base G dans une région polyG localisée dans l'exon 3 de *FT-D* (*FT-D<sub>ex3</sub>*, noté FTD (G)<sub>3or4</sub> dans Bonnin *et al.* 2008).

Ces trois polymorphismes ont été révélés au cours d'un travail préliminaire de recherche de gènes candidats réalisé dans l'équipe (Bonnin *et al.* 2008). Le génotypage a été réalisé par séquençage dans les locaux du CNG (Centre National de Génotypage, Evry, 91 France) à la plateforme INRA.

### II-4.4 Génotypage de *LD*, *CO* et *GI*

La recherche de polymorphisme dans des séquences orthologues de gènes impliqués dans la floraison chez *Arabidopsis thaliana* et le riz menée par Isabelle Bonnin en préambule de ce travail a conduit à l'identification de quatre sites localisés dans 3 gènes potentiellement impliqués dans la variation pour la précocité de floraison chez le blé et polymorphes chez les 16 lignées parentales de GD :

- Séquence orthologue du gène *LUMINIDEPENDENS* (*LD*) : 2 SNP (polymorphismes C/T et A/C) localisés dans l'intron de blé équivalent à l'intron 11 de *LD* du riz et sur la copie du génome B du blé (notés *LD-B<sub>int11a</sub>* et *LD-B<sub>int11b</sub>* dans la suite) et séparés de 310 pb l'un de l'autre.

- Séquence orthologue du gène *CONSTANS* (*CO*) : un SNP (polymorphisme A/G) localisé dans l'intron équivalent de l'intron de *CO* du riz et sur la copie du génome B du blé (*CO-B<sub>int</sub>*). *CONSTANS* constitue en fait une famille de gènes, l'équivalent du gène chez le riz (nommé Hd1) a été identifié mais cette copie du gène est monomorphe chez les 16 parents de la population de GD. La copie de *CONSTANS* amplifiée ici est en fait un équivalent du gène identifié chez l'orge (HvCO3).

- Séquence orthologue du gène *GIGANTEA* (*GI*) : un SNP (polymorphisme A/C) localisé dans l'intron 4 de blé équivalent de l'intron 4 de *GI* du riz et sur la copie du génome B du blé (*GI-B<sub>ex5</sub>*).

Du fait de la nature hexaploïde du génome du blé, le génotypage a été effectué en 2 étapes :

(i) Amplification spécifique de la copie du génome (A, B ou D) à amplifier en se basant sur du polymorphisme trouvé entre copies homéologues du gène. Cette étape nécessite la mise au point de conditions d'amplification génome spécifique et la vérification de la spécificité en amplifiant simultanément la séquence chez le blé hexaploïde avec la séquence chez les blés diploïdes ancêtres (*Triticum urartu*, ancêtre du génome A, *Triticum speltoides*, ancêtre supposé du génome B et *T. tauschii*, ancêtre du génome D, Figure II-9) ;

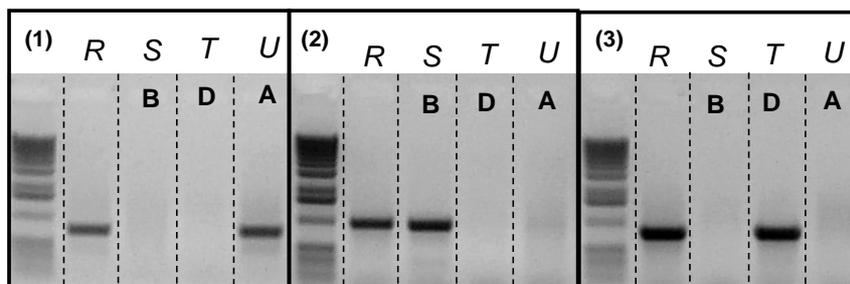


Figure II-9 : Exemple d'amplification génome-spécifique de fragments du génome A (1) du génome B (2) et du génome D (3). R= Récital, variété de blé hexaploïde ; S= *T. speltoides*, espèce diploïde porteuse du génome B, T= *T. tauschii*, espèce diploïde porteuse du génome D et U= *T. urartu*, espèce diploïde porteuse du génome A.

(ii) Génotypage du site polymorphe proprement dit. L'ensemble de ces polymorphismes a été génotypé selon une technique dite de PCR allèle-compétitive développée à l'INRA du Moulon par l'équipe BMBI et qui permet de génotyper des SNP de façon co-dominante à moindre coût

en utilisant l'électrophorèse sur gel d'agarose. Le génotypage des polymorphismes a été réalisé en partie par l'équipe de BMBI (Figure II-10).

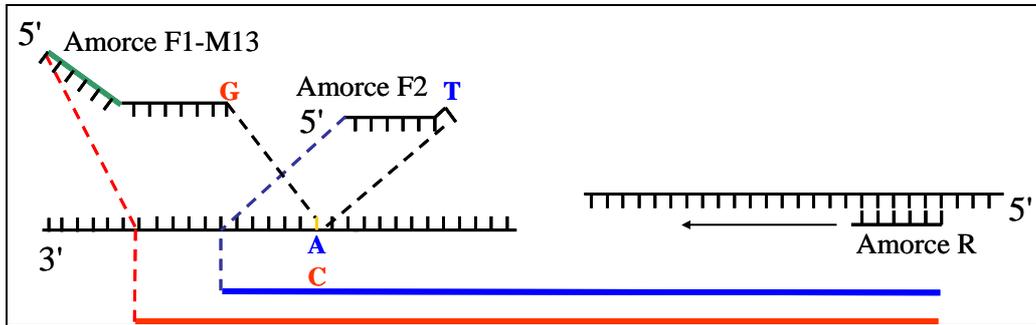


Figure II-10 : Principe de la PCR allèle compétitive : 3 amorces sont introduites dans le mélange de réaction PCR : une amorce « Reverse » (amorce R) et 2 amorces « Forward » (amorce F) possédant en position 3' la base complémentaire des 2 bases observées au site polymorphe. Une des 2 amorces est prolongée par une queue M13 longue de 20pb en 5' (Amorce F1-M13). Les produits d'amplification obtenus à l'issue de la PCR présentent donc un polymorphisme de taille (en rouge et bleu) en fonction de la base présente au site polymorphe et sont différenciables par migration sur gel d'agarose à 4%. (D'après M. Falque, équipe BMBI, INRA du Moulon).

## *Chapitre 3*

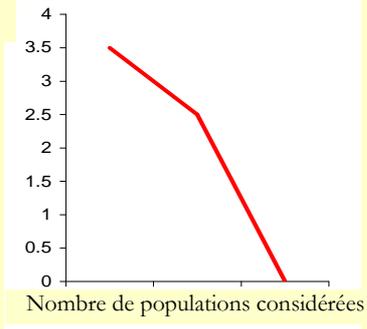
# *SUIVI DE L'ÉVOLUTION DES MARQUEURS NEUTRES DU GENOME DANS LES POPULATIONS DE GESTION DYNAMIQUE*

# Chapitre III : Suivi de l'évolution des marqueurs neutres du génome dans les populations de gestion dynamique

## III-1 Introduction

Des populations évoluant en conditions naturelles sont soumises à la fois à des pressions de sélection affectant des locus directement, ou indirectement, impliqués dans la valeur sélective des individus et à des forces affectant aléatoirement la diversité neutre qui est fixée de la même façon sur l'ensemble du génome. Dans toute population de taille finie, les allèles présents à une génération représentent un échantillon des allèles présents à la génération précédente. Cet échantillonnage entraîne des variations aléatoires de fréquences alléliques d'une génération à l'autre, dont la conséquence à plus ou moins long terme est la perte aléatoire d'allèles à divers locus. Ce mécanisme dit de dérive génétique affecte l'ensemble du génome et dépend en particulier de la taille de la population et du régime de reproduction. La perte de diversité peut être compensée à long terme par des événements de mutation, et dans le cas de populations non isolées, par l'arrivée d'individus migrants susceptibles d'apporter de la diversité à l'ensemble du génome. Lorsque l'on cherche à comprendre l'histoire évolutive de populations, l'estimation de la dérive et l'identification des événements démographiques subis par chacune des populations est une étape importante car ces facteurs peuvent influencer la réponse à la sélection ou même mimer les effets de la sélection elle-même (Stinchcombe et Hoekstra, 2007). Par exemple, des populations de faible effectif vont subir des effets importants de dérive pouvant entraîner de fortes variations de fréquences alléliques et conduire à la fixation rapide d'allèles favorables, neutres ou même faiblement délétères, du simple fait du hasard. Dans un tel contexte, les effets démographiques et sélectifs peuvent être difficiles à dissocier et la sélection pourra avoir moins d'effet sur les gènes impliqués dans la valeur sélective (Glémin *et al.* 2006). Dans le cas des populations de blés de gestion dynamique, les populations sont majoritairement autogames (reproduction par autofécondation) et évoluent de façon isolée, du fait du dispositif mis en place. L'impact de l'autogamie sur la diversité génétique des populations a été longuement commenté dans un article récent (Charlesworth 2003). Un des effets majeurs du régime de reproduction est la réduction des effectifs efficaces dans les populations autogames en comparaison des

**Encadré III-1 : Principaux résultats des travaux antérieurs portant sur l'évolution de la diversité neutre de populations du réseau de gestion dynamique.**

	<b>Thèse J. Enjalbert (1999)</b>	<b>Thèse AL. Raquin (2005)</b>									
<b>Type de marqueurs</b>	RFLP (15 marqueurs amplifiant 23 locus)	Microsatellites (20 marqueurs amplifiant 21 locus (1 par chromosome))									
<b>Populations et générations étudiées</b>	<ul style="list-style-type: none"> <li>• Populations de Rennes, Le Moulon, Toulouse</li> <li>• Population initiale (G0) et Génération 10</li> <li>• ≈ 80 individus par populations étudiées</li> </ul>	<ul style="list-style-type: none"> <li>• Population du Moulon</li> <li>• Générations 1, 5, 10, 16, 17</li> <li>• 110 à 330 individus par générations</li> </ul>									
<b>Régime de reproduction</b>	<ul style="list-style-type: none"> <li>• Régime de reproduction majoritairement autogame</li> <li>• Taux d'allofécondation résiduel variable selon les populations en génération 10</li> </ul> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td>Le Moulon</td> <td>Rennes</td> <td>Toulouse</td> </tr> <tr> <td>Taux d'allofécondation</td> <td>5,9%</td> <td>6.7%</td> <td>10.1%</td> </tr> </table>		Le Moulon	Rennes	Toulouse	Taux d'allofécondation	5,9%	6.7%	10.1%	<ul style="list-style-type: none"> <li>• Taux d'allofécondation résiduel stable après 17 générations d'évolution dans la population du Moulon (<math>t=3.1\%</math>)</li> </ul>	
	Le Moulon	Rennes	Toulouse								
Taux d'allofécondation	5,9%	6.7%	10.1%								
<b>Effectifs Efficaces</b>	<ul style="list-style-type: none"> <li>• Les faibles effectifs efficaces (<math>N_e</math>) estimés à partir des variations temporelles de fréquences alléliques entre 2 générations dans les différentes populations suggèrent que les populations subissent d'importantes pressions de dérive et de sélection</li> </ul> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td>Le Moulon</td> <td>Rennes</td> <td>Toulouse</td> </tr> <tr> <td><math>N_e</math> estimés entre G0 et G10 [CI95%]</td> <td>123 [58–277]</td> <td>56 [29–103]</td> <td>187 [81–543]</td> </tr> </table>		Le Moulon	Rennes	Toulouse	$N_e$ estimés entre G0 et G10 [CI95%]	123 [58–277]	56 [29–103]	187 [81–543]	<ul style="list-style-type: none"> <li>• Effectif efficace estimé à partir des variations temporelles de fréquences alléliques entre la population parentale et la génération 17 de la population du Moulon : <math>N_e=167</math>, CI95% = [119–216]</li> <li>• Cette valeur, comparable à la valeur estimée après 10 générations d'évolution (Enjalbert 1999), suggère que la population est soumise à une intensité de dérive et de sélection constante au cours des générations.</li> </ul>	
	Le Moulon	Rennes	Toulouse								
$N_e$ estimés entre G0 et G10 [CI95%]	123 [58–277]	56 [29–103]	187 [81–543]								
<b>Structuration de la diversité</b>	<ul style="list-style-type: none"> <li>• Des allèles sont perdus dans différentes populations du fait de la dérive, mais le maintien de la diversité à l'échelle du réseau est assuré par le dispositif en multi populations</li> <li>• Nombre moyen d'allèles perdus</li> </ul>  <p style="text-align: center;">Nombre de populations considérées</p> <ul style="list-style-type: none"> <li>• Les populations se différencient sous l'effet combiné de la dérive et de la sélection</li> </ul> <p><u>Différenciation multilocus à la génération 10:</u></p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td>Le Moulon</td> <td>Toulouse</td> </tr> <tr> <td>Toulouse</td> <td>0.03</td> <td style="background-color: #cccccc;"></td> </tr> <tr> <td>Rennes</td> <td>0.04</td> <td>0.05</td> </tr> </table> <p style="text-align: center;">(source : Enjalbert <i>et al.</i> 1999)</p>		Le Moulon	Toulouse	Toulouse	0.03		Rennes	0.04	0.05	<ul style="list-style-type: none"> <li>• Faible diminution de la diversité microsatellite intra-population après 16 générations d'évolution : la diversité de <math>N_e</math>, <math>H_e</math>, passe de 0.57 dans la population parentale à 0.53 en génération 17.</li> <li>• Apparition d'une nouvelle diversité aux locus microsatellites du fait du fort taux de mutation de ce type de marqueurs (taux de mutation de 0 à <math>5.10^{-3}</math> selon les locus)</li> </ul>
	Le Moulon	Toulouse									
Toulouse	0.03										
Rennes	0.04	0.05									

populations allogames. L'effectif efficace (noté  $N_e$ ) est défini comme la taille d'une population idéale (population isolée, de taille constante, se reproduisant au hasard et pour laquelle le nombre de descendants par parent suit une loi de Poisson (Wright 1931) qui présenterait le même niveau de dérive que la population étudiée. L'effectif efficace peut également être vu comme le nombre d'individus de la génération précédente ayant contribué à la formation de la génération observée. Un fort niveau d'homozygotie est attendu à chaque locus dans les populations autogames, ce qui tend à réduire de moitié le nombre de gamètes différents disponibles pour constituer la génération suivante par rapport aux populations allogames de même taille démographique. Un faible effectif efficace se traduit par un fort effet de dérive affectant l'ensemble du génome et donc, à terme, une perte de diversité génétique. Dans le cas des populations de gestion dynamique, le risque de perte de diversité intra population est accru par l'isolement des populations imposé par le dispositif expérimental. Une certaine différenciation génétique entre les populations est attendue du simple fait de la dérive conduisant à la fixation au hasard d'allèles différents dans les différentes populations étudiées. Ainsi on s'attend à ce que les populations de gestion dynamique de blé comme, et même plus généralement la plupart des populations de plantes annuelles autogames, montrent un niveau de variabilité intra population réduit mais un niveau de différenciation entre populations élevé (Charlesworth *et al.* 1997). Ce patron de diversité étant également celui attendu pour les caractères ou les gènes soumis à sélection lors de l'adaptation locale (Le Corre 2005), la sélection dans chacune des populations étudiées pourra être difficile à dissocier des effets liés à la dérive et à la démographie. L'objectif de ce chapitre est donc de suivre l'évolution de la diversité de locus *a priori* neutres répartis sur l'ensemble du génome afin d'identifier les forces qui affectent l'ensemble du génome dans les populations étudiées ici. Ceci nous permettra de mieux dissocier par la suite les facteurs purement démographiques des effets de la sélection locus-spécifique et de mieux appréhender la capacité de réponse à la sélection des populations étudiées et donc leur capacité d'adaptation.

Les marqueurs neutres et co-dominants, tels que les marqueurs microsatellites, sont des outils intéressants pour appréhender les pressions évolutives affectant l'ensemble du génome et caractériser le régime de reproduction. Deux études antérieures portant sur des populations différentes ou échantillonnées à des périodes différentes de celles utilisées dans ce travail (Enjalbert 1999 ; Raquin 2005 ; *cf.* Encadré III-1) ont permis de montrer que le régime de reproduction de ces populations est majoritairement autogame. De plus, malgré les faibles effectifs efficaces estimés dans ces populations, (i) la diversité neutre est maintenue à

l'échelle du réseau pour les marqueurs RFLP après 10 générations d'évolution et une nouvelle diversité apparaît aux marqueurs microsatellites après 16 générations d'évolution du fait du fort taux de mutation observé à ces marqueurs et (ii) les populations évoluant sur des sites différents se sont différenciées après 10 générations d'évolution permettant un maintien de la diversité génétique à l'échelle du réseau. L'utilisation de marqueurs microsatellites est discutable pour étudier l'évolution de la diversité sur l'ensemble du génome du fait du fort taux de mutation observé pour ce type de marqueurs (Raquin 2005) et du grand nombre d'allèles qui en résulte rendant complexe l'étude, par exemple, du déséquilibre de liaison. Cependant ce type de marqueurs a été intensivement développé chez le blé et un grand nombre de marqueurs et de cartes génétiques sont disponibles. Ainsi, dans ce chapitre, l'évolution de la diversité neutre est étudiée dans les lignées parentales, la population initiale G0 et les 3 populations de GD échantillonnées à 3 générations en se basant sur les 21 marqueurs microsatellites disponibles, répartis sur l'ensemble du génome (un par chromosome, cf. Figure II-7). Une étude plus fine de l'étendue et l'évolution du déséquilibre de liaison est également menée sur 7 marqueurs microsatellites répartis le long du chromosome 3B. Les effectifs efficaces, le régime de reproduction et la structuration de la diversité sont estimés de façon spécifique dans chacune des populations et une approche multilocus originale nous permet de suivre l'évolution des associations alléliques qui se mettent en place au cours des générations.

## III-2. Méthodes

### III-2.1 Diversité

Les fréquences alléliques, le nombre moyen d'allèles par locus et par populations et la diversité de Nei ont été estimés à l'aide du GENETIX v4.05.2 (Belkhir *et al.* 2004). La diversité génétique intra population a été estimée pour chaque locus à chaque génération par

$$H_e = \frac{N}{N-1} \left( 1 - \sum_{i=1}^n \hat{p}_i^2 \right) \quad (\text{Nei, 1987})$$

$N$  étant le nombre d'individus échantillonnés dans la population considérée,  $n$  le nombre d'allèles au locus étudié et  $\hat{p}_i$  la fréquence estimée de l'allèle  $i$ . Cet estimateur, appelé couramment « diversité de Nei », mesure la fréquence d'hétérozygotes attendus dans le cas d'une population idéale panmictique (croisement au hasard des gamètes), mais il est aussi égal à la probabilité de tirer au hasard 2 gamètes non identiques dans une population. Calculée à l'échelle d'un locus, cette mesure est très dépendante du nombre d'allèles à ce même locus et donc aussi du type de marqueurs utilisé. La diversité moyenne par population a ensuite été estimée sur les 21 marqueurs microsatellites. La correction pour le biais d'échantillonnage ( $N/(N-1)$ ) n'a pas été appliquée pour estimer la diversité des 16 lignées parentales qui ont toutes été génotypées.

### III-2.2 Taux d'allo-fécondation

Le taux d'allofécondation a été dans un premier temps estimé en se basant sur le coefficient de consanguinité intra population,  $F_{IS}$ , mesurant le déficit en hétérozygotes par rapport à la valeur attendue sous l'hypothèse de croisement au hasard des gamètes :

$$F_{IS} = 1 - \frac{H_o}{H_e}$$

avec  $H_o$  et  $H_e$ , respectivement fréquences observées et attendues d'hétérozygotes. Cette valeur, ainsi estimée, correspond également à la probabilité que 2 allèles observés à un locus soient identiques « par descendance ». La valeur moyenne des  $F_{IS}$  estimés sur les 21 locus

microsatellites a été utilisée pour estimer le taux d'allofécondation  $t$  à partir de la relation entre  $F_{IS}$  et  $s$  (taux d'autofécondation,  $s = 1 - t$ ) :

$$F_{IS} = \frac{s}{2 - s}$$

Cette relation suppose que le taux d'alfofécondation a été constant pendant un grand nombre de générations, que la population est à l'équilibre pour le régime de reproduction et que l'autofécondation est seule responsable de la déviation des fréquences génotypiques par rapport aux proportions de Hardy-Weinberg (*i.e.* en l'absence de structuration spatiale intra population et de différences de valeur sélective entre descendants issus d'alfofécondation et descendants issus d'autofécondation).

Comme la méthode basée sur les  $F_{IS}$  est sensible aux erreurs de génotypage pouvant conduire à une sous-estimation du taux d'alfofécondation, nous avons également utilisé une approche multi-locus d'estimation du taux d'autofécondation récemment proposée par David *et al.* (2007). Cette approche se base sur la corrélation d'hétérozygotie à plusieurs locus et donc limite les effets des erreurs de génotypage affectant *a priori* chaque locus indépendamment des autres. Les individus d'une population fortement autogame tendent à avoir la plupart de leur locus à l'état homozygote créant ainsi un déséquilibre d'identité. Un évènement d'alfofécondation va conduire à la présence de plusieurs locus à l'état hétérozygote simultanément chez le même individu. Ainsi dans ce cas, la distribution de l'hétérozygotie multi-locus dévie de la distribution attendue sous l'hypothèse de répartition au hasard des locus hétérozygotes. Deux méthodes ont été proposées en se basant sur cette approche :

- L'une se fonde sur le déséquilibre d'hétérozygotie multilocus à partir d'une estimation combinant le déséquilibre estimé entre toutes les autres paires de locus ( $g_2$ ). Cet estimateur est ensuite utilisé pour estimer le taux d'autofécondation  $s$ .
- L'autre estimation est obtenue par maximum de vraisemblance en choisissant la valeur de  $s$  qui maximise la probabilité d'obtenir la distribution d'hétérozygotie multilocus apparente observée.

Les 2 méthodes produisent des résultats équivalents (David *et al.* 2007), et seule la première méthode est présentée ici pour estimer les taux d'autofécondation dans les populations à l'aide du programme RMES (Robust Multilocus Estimation of Selfing) développé par David *et al.* (2007). Ces valeurs ont été converties en taux d'alfofécondation  $t$  ( $t = 1 - s$ ) afin d'établir leur lien avec les études précédemment menées sur les populations de gestion dynamique.

### III-2.3 Structuration spatiale et temporelle

Dans un premier temps, une analyse factorielle des correspondances (AFC) a été menée pour estimer les distances génétiques entre populations et évaluer leur niveau de différenciation. Cette analyse consiste à projeter les distances génétiques entre individus sur un jeu d'axes qui maximisent les différences entre populations. Elle a été effectuée à l'aide du logiciel GENETIX v4.05.2 (Belkhir *et al.* 2004).

Dans le cas de populations génétiquement divergentes, un paramètre de différenciation entre populations peut être estimé. Il mesure l'écart entre le taux d'hétérozygotes attendu globalement si les populations échangent librement des gènes (panmixie) et le taux d'hétérozygotes moyen observé dans les populations. Cette mesure peut également être vue comme la proportion de diversité totale expliquée par l'écart des fréquences alléliques entre populations. Dans notre matériel, la différenciation spatiale estimée de cette façon est un indicateur de diversité à l'échelle du réseau de populations de gestion dynamique. C'est l'estimateur  $\theta$  du  $F_{ST}$ , proposé par Weir et Cockerham (1984) qui a été calculé ici grâce au logiciel FSTAT version 2.9.3.2 (Goudet 2002).

$$\theta = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma_w^2}$$

avec  $\sigma_a^2$  variance inter populations,  $\sigma_b^2$  variance intra population, et  $\sigma_w^2$  variance intra individu.

Pour tester si les valeurs de  $\theta$  observées entre populations à chaque génération sont statistiquement différentes de zéro, des permutations de génotypes entre les populations échantillonnées ont été réalisées avec FSTAT.

### III-2.4 Variations de fréquences alléliques et effectifs efficaces

L'évolution temporelle des fréquences alléliques a été suivie en estimant à chaque locus la variance temporelle standardisée des fréquences alléliques ( $\hat{F}_{c,l}$ ) entre les différentes générations étudiées :

$$\hat{F}_{c,l} = \frac{1}{K} \sum_{k=1}^K \frac{(p_{ki} - p_{kf})^2}{\left( \frac{(p_{ki} + p_{kf})}{2} - p_{ki} \times p_{kf} \right)} \quad (\text{Nei et Tajima, 1981})$$

$K$  étant le nombre d'allèles au locus  $l$ , et  $p_{ki}$  et  $p_{kf}$ , respectivement, les fréquences alléliques dans les échantillons de la population initiale et finale. Le  $\hat{F}_c$  moyen, égal à la moyenne des  $\hat{F}_{c,l}$  pondérée par le nombre d'allèles à chaque locus, a ensuite été calculé :

$$\hat{F}_c = \frac{\sum_l (K_l \hat{F}_{c,l})}{\sum_l K_l}$$

Cet estimateur de la variance standardisée des fréquences alléliques permet de calculer l'effectif efficace génétique ( $Ne$ ) entre les différentes générations. L'effectif efficace correspond ici à la taille d'une population idéale panmictique qui, sous le seul effet de la dérive, génèrerait les mêmes variations de fréquences alléliques que celles observées dans la population étudiée. Puisque les individus échantillonnés ne participent pas à la reproduction (sampling scheme II, Waples, 1989), il est estimé sous l'hypothèse qu'il n'y a pas de corrélation entre les fréquences alléliques estimées et l'effectif efficace :

$$\hat{N}e = \frac{t}{2 \left( Fc - \frac{1}{S_0} - \frac{1}{S_t} \right)}$$

où  $t$  est le nombre de générations séparant les 2 populations étudiées, et  $S_0$  et  $S_t$ , respectivement, la taille de l'échantillon dans la population initiale et dans la population finale (cas autogame). Les intervalles de confiance des  $Ne$  ont été calculés en se basant sur le fait que  $nFc/E(Fc)$  suit une loi de distribution pouvant être approximée par un chi2 à  $n$  degrés de liberté (Lewontin et Krakauer, 1973) ( $n = \sum_l (K_l - 1)$  avec  $K_l$  nombre d'allèles au locus  $l$ ). L'estimation de l'effectif efficace présentée ici peut conduire à des valeurs de  $Ne$  qui tendent vers l'infini lorsque la variance d'échantillonnage approche la variance liée à la dérive ( $1/S_0 - 1/S_t \approx Fc$ ) ou à des valeurs négatives de  $Ne$  lorsque la variance d'échantillonnage est supérieure à la variance liée à la dérive ( $1/S_0 - 1/S_t > Fc$ ). Aucune valeur négative n'a été

observée pour les estimations de  $Ne$  à partir du  $\hat{F}_c$  moyen, mais des bornes supérieures extrêmes ou négatives ont été trouvées lors de l'estimation de l'intervalle de confiance. Dans ce cas, la valeur de la borne de  $Ne$  a été remplacée par l'effectif démographique (au moins 5000 graines re-semées chaque année) corrigé par le coefficient de consanguinité ( $Ne_d = N/(1+Fis) \approx 2650$ ).

### III-2.5 Déséquilibre de liaison

Le déséquilibre de liaison (DL) entre locus pris 2 à 2 a été mesuré par des coefficients de corrélations ( $R$ , cf. Encadré III-2). Ces coefficients de corrélation entre les 2 allèles de chaque paire et le coefficient de corrélation moyen par paire de locus ont été calculés à l'aide du programme LINKDIS (Garnier-Gere et Dillmann, 1992), disponible avec le logiciel GENETIX, qui se base sur l'estimation du déséquilibre de liaison calculé entre paires d'allèles proposé par Weir (1979) :

$$R_{ij} = \frac{\Delta_{ij}}{\left[ \left( p_i(1-p_i) + D_{i \cdot}^i \right) \left( q_j(1-q_j) + D_{\cdot j}^j \right) \right]^{\frac{1}{2}}}, \quad \text{avec} \quad \Delta_{ij} = \frac{N}{N-1} \left( \frac{T_{ij}}{N} - 2p_i q_j \right)$$

$T_{ij}$  étant le nombre de fois où l'allèle  $i$  d'un locus apparaît associé avec l'allèle  $j$  d'un autre locus dans le même individu, et  $p_i$  et  $q_j$  étant respectivement la fréquence de l'allèle  $i$  et de l'allèle  $j$  dans la population échantillonnée de  $N$  individus. Le coefficient de corrélation  $R_{ij}$  est ici corrigé par  $\Delta_j$  qui prend en compte l'écart à la panmixie (Black et Krafur 1985).

Comme les classes d'allèles sous-représentées sont susceptibles de biaiser l'estimation du DL, les corrélations ont été estimées entre les allèles présents à une fréquence supérieure à 5% dans la population considérée. Les coefficients de corrélation moyens pour chaque paire de locus ont ensuite été estimés par la moyenne empirique des  $R_{ij}$  estimés dans chacune des populations.

Pour tester si les corrélations observées intra paires de locus sont significativement différentes de zéro, un test exact de déséquilibre de liaison gamétique a été effectué dans chaque population entre paires de locus à l'aide du programme Genepop v4 (Raymond et Rousset, 1995) en testant l'hypothèse nulle « le génotype à un locus est indépendant du génotype à un

autre locus ». Les déséquilibres de liaison ont également été estimés et testés pour les 7 locus microsatellites répartis le long du chromosome 3B.

### III-2.6 Reconstruction d'haplotypes multilocus

Du fait des forts déséquilibres de liaison observés entre les 21 locus étudiés, des génotypes multilocus ont été reconstitués. Par la suite, on appellera « haplotype » l'association des allèles des différents locus à l'état haploïde. Lors de la reconstruction, les individus présentant des données incomplètes sur le jeu de locus étudiés (données manquantes à au moins l'un des locus) ont été supprimés. Pour ne pas trop diminuer la taille des échantillons de populations analysés, les locus présentant plus de 8% de données manquantes (5 locus au total) n'ont pas été considérés pour reconstruire les haplotypes. De même, 2 locus présentant un niveau de diversité faible dans l'ensemble des populations (*Xgwm480* et *Xgwm664*, locus bi-alléliques, *He* compris entre 0 et 0,18 ; Tableau III-1) n'ont pas été considérés pour reconstruire les haplotypes car ces locus sont peu informatifs. Ainsi la reconstruction des haplotypes présentée ici porte sur 14 locus répartis sur des chromosomes différents et a été effectuée à l'aide du logiciel ARLEQUIN v3.01 (Excoffier *et al.* 2005) en utilisant l'algorithme ELB de reconstruction d'haplotypes basé sur une approche bayésienne. La reconstruction d'haplotypes est délicate dans le cas de génotypes présentant plus de 2 locus (sur les 14 considérés) à l'état hétérozygote. Pour ces individus, l'algorithme utilisé permet de retenir les associations alléliques les plus probables en comparaison des associations alléliques observées sur l'ensemble du jeu de données. La reconstruction d'haplotypes a également été effectuée sur le jeu de 7 locus microsatellites répartis le long du chromosome 3B.

### Encadré III-2 : Définition et mesure du déséquilibre de liaison (DL), d'après Ardlie et al. 2002.

**Définition :** Déséquilibre de liaison : Ecart à l'association aléatoire d'allèles à 2 locus différents.

Cas de 2 locus bialléliques :

Locus 1, présentant les allèles A et a aux fréquences  $p_A$  et  $p_a$

Locus 2, présentant les allèles B et b aux fréquences  $p_B$  et  $p_b$

→ Les 2 locus sont indépendants : fréquence de l'haplotype  $p_{AB} = p_A \times p_B$

→ Les 2 locus sont en déséquilibre de liaison :  $p_{AB} \neq p_A \times p_B$

**Mesures :**

- Ecart entre la fréquence attendue dans le cas de l'indépendance des 2 locus et la fréquence observée de l'haplotype :

$$D = p_{AB} - p_A \times p_B$$

L'estimation du DL par la mesure de D est fortement dépendante des fréquences alléliques à chacun des locus. Pour s'affranchir de cette dépendance et pouvoir comparer le DL entre paires de locus différentes, 2 principales mesures sont couramment utilisées :

Déséquilibre de liaison normalisé :  $D'$

$$D' = \frac{D}{D_{MAX}} \text{ avec } D_{MAX} = \text{Min}[p_A p_B, p_a p_b] \text{ si } D < 0$$
$$D_{MAX} = \text{Min}[p_a p_B, p_A p_b] \text{ si } D > 0$$

$D'=1$  si 2 locus ne sont séparés par aucun évènement de recombinaison. Cependant, l'interprétation de  $D'<1$  est peu claire, la valeur de  $D'$  dépend fortement de la taille de l'échantillon considéré ce qui rend impossible son utilisation pour comparer des échantillons de tailles différentes. De plus, cette méthode n'est pas totalement indépendante des fréquences alléliques et tend à accorder un poids plus fort aux allèles rares.

Coefficient de corrélation entre allèles :  $r^2$  ou  $r$

$$r^2 = \frac{D^2}{p_A p_a p_B p_b}$$

$r^2 = 1$  si 2 locus ne sont séparés par aucun évènement de recombinaison et ont les mêmes fréquences alléliques. Les valeurs de  $r^2$  donnent le niveau d'information fourni par un locus sur un autre locus. Cette mesure permet de quantifier et de comparer des mesures de DL entre plusieurs locus.

## III-3 Résultats

### III-3.1 Diversité génétique, apparition de nouveaux allèles

Comme attendu dans le cas de marqueurs microsatellites, le niveau de polymorphisme initial est très important dans la population parentale avec un total de 81 allèles sur les 21 locus microsatellites étudiés, soit un nombre moyen d'allèles par locus de 3,85. Dix-sept nouveaux allèles, non présents chez les parents sont recensés dans la population initiale G0. Tous ces allèles sauf un sont retrouvés dans au moins une des populations étudiées aux différentes générations. D'autre part, 20 nouveaux allèles apparaissent au cours des générations dans les différentes populations. Parmi ces allèles, 3 sont présents dans au moins 2 des populations étudiées dès les premières générations, ce qui laisse penser qu'ils étaient déjà présents dans la population initiale mais qu'ils n'ont pas été échantillonnés dans notre matériel. Cependant la plupart des autres nouveaux allèles apparaissent de façon spécifique dans les différentes populations : 4 allèles spécifiques de la population du Moulon, 3 allèles spécifiques de Toulouse et 7 allèles spécifiques de Vervins. Deux allèles présents dans les lignées parentales à 2 locus différents ne sont ré-échantillonnés dans aucune des 3 populations étudiées en génération 12. La diversité de Nei ( $H_e$ ) estimée dans les populations de Toulouse et Le Moulon est du même ordre de grandeur que celle estimée dans la population parentale et dans la population initiale G0 et évolue peu au cours des générations. Dans le cas de la population de Vervins, le niveau de diversité est réduit dès la première génération étudiée (G2) comparée à celle de G0 (Tableau III-2).

Tableau III-2 : Diversité de Nei et nombre moyen d'allèles par locus estimés dans chacune des populations à chaque génération.

Population	Génération	$H_e$	Ecart type	Nombre moyen d'allèles par locus
P. parentale		0.55	0.19	3.95
P. initiale	G0	0.56	0.20	4.24
Le Moulon	G2	0.52	0.21	4.38
	G7	0.53	0.19	3.90
	G12	0.53	0.22	4.19
Toulouse	G2	0.54	0.20	4.43
	G7	0.54	0.20	4.14
	G12	0.53	0.21	3.76
Vervins	G2	0.45	0.22	4.24
	G7	0.49	0.21	4.14
	G12	0.42	0.21	3.67

Tableau III-1

Tableau III-1 : Détail des estimations par locus des paramètres de diversité : nombre d'allèles, diversité de Nei ( $H_e$ ) et coefficient de consanguinité ( $F_{is}$ ) par population.

Locus	Xqum135	Xqum141	Xqum42	Xqum172	Xqum257	Xqum261	Xqum480	Xqum566	Xqum64	Xqum77	Xqum149	Xqum149	Xqum115	Xqum234	Xqum272	Xqum427	Xqum626	Xqum225	Xqum260	Xqum400	Xqum428
Position	1A	1B	1D	2A	2B	2D	3A	3B	3D	4A	4B	4D	5A	5B	5D	6A	6B	6D	7A	7B	7D
Nombre d'allèles parentaux	4	4	3	5	2	4	2	5	2	2	3	6	3	6	4	6	3	4	6	5	4
Nombre total de nouveaux allèles	0	4	1	4	1	1	2	1	1	1	0	5	1	2	2	1	2	2	1	0	1
Nombre total Le Moulon	2	5	4	6	2	4	3	6	2	2	2	7	3	6	5	6	3	6	6	5	4
G12	4	4	4	4	2	4	2	6	1	2	3	4	6	3	6	6	6	6	6	4	3
Toulouse	2	3	2	7	2	3	2	4	2	3	3	6	3	6	5	4	2	5	4	6	3
Vervins	2	3	2	7	2	3	2	4	2	3	3	6	3	6	5	4	2	5	4	6	3
P. parentale	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
P. initiale	0.49	0.33	0.62	0.66	0.49	0.58	0.12	0.75	0.12	0.30	0.54	0.70	0.54	0.63	0.58	0.77	0.46	0.60	0.70	0.67	0.60
G30	31	35	34	34	34	35	36	31	34	31	34	36	34	36	32	34	34	35	31	33	36
Le Moulon	0.48	0.40	0.60	0.67	0.51	0.70	0.08	0.64	0.11	0.36	0.63	0.73	0.49	0.70	0.72	0.82	0.44	0.69	0.66	0.78	0.58
G12	0.87	1	1	0.96	1	0.96	0.66	0.95	1	1	1	0.96	1	1	1	0.97	1	1	0.90	1	1
P. parentale	56	59	57	54	48	50	53	56	57	55	54	57	56	55	53	54	56	46	56	55	57
P. initiale	0.27	0.36	0.52	0.69	0.47	0.56	0.04	0.69	0.03	0.36	0.53	0.72	0.55	0.71	0.60	0.81	0.46	0.67	0.57	0.73	0.59
G30	0.93	0.95	0.87	0.95	0.91	0.96	1	0.97	1	0.90	0.90	0.90	0.90	0.80	0.88	0.98	1	0.94	0.88	0.98	0.94
Le Moulon	0.34	0.47	0.56	0.54	0.51	0.55	0.18	0.74	0.00	0.36	0.49	0.67	0.64	0.57	0.66	0.73	0.44	0.68	0.67	0.73	0.64
G12	0.95	1	0.94	0.94	0.93	0.97	1	1	0.90	1	0.92	1	0.91	0.91	0.89	0.98	1	0.98	1	0.95	0.95
P. parentale	54	58	59	58	59	60	59	57	58	55	60	59	56	59	56	60	60	60	60	59	59
P. initiale	0.18	0.55	0.67	0.70	0.41	0.64	0.05	0.67	0.07	0.36	0.36	0.72	0.62	0.67	0.52	0.80	0.40	0.76	0.67	0.75	0.66
G30	0.90	0.94	0.98	0.95	1	0.97	0.67	0.97	1	0.90	0.91	0.91	0.97	0.98	0.83	0.98	1	0.98	0.97	1	1
Le Moulon	0.34	0.47	0.56	0.54	0.51	0.55	0.18	0.74	0.00	0.36	0.49	0.67	0.64	0.57	0.66	0.73	0.44	0.68	0.67	0.73	0.64
G12	0.95	1	0.94	0.94	0.93	0.97	1	1	0.90	1	0.92	1	0.91	0.91	0.89	0.98	1	0.98	1	0.95	0.95
P. parentale	52	58	59	57	60	54	59	55	60	57	57	58	59	60	47	60	60	60	60	58	54
P. initiale	0.58	0.47	0.67	0.63	0.51	0.61	0.13	0.75	0.03	0.29	0.54	0.62	0.55	0.72	0.52	0.77	0.45	0.71	0.74	0.68	0.38
Toulouse	1	1	0.98	0.86	0.84	0.97	1	0.98	1	1	0.94	0.95	0.91	0.98	1	0.91	0.96	0.93	0.98	0.98	1
Le Moulon	0.40	0.33	0.64	0.57	0.48	0.71	0.18	0.75	0.00	0.43	0.58	0.63	0.49	0.73	0.47	0.80	0.48	0.70	0.75	0.67	0.58
G12	0.87	1	0.97	1	0.93	0.98	1	0.93	1	0.96	1	0.97	1	1	1	0.95	1	0.97	0.93	0.95	1
P. parentale	57	59	58	59	60	59	55	56	56	58	59	58	55	59	53	60	57	59	59	59	56
P. initiale	0.36	0.38	0.56	0.64	0.50	0.69	0.07	0.62	0.00	0.49	0.59	0.67	0.44	0.74	0.45	0.80	0.49	0.73	0.72	0.66	0.59
Vervins	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.95	0.98	1	0.97
G2	56	59	58	57	59	53	52	56	59	60	60	60	57	57	58	60	57	58	57	60	59
P. parentale	0.41	0.32	0.28	0.72	0.16	0.57	0.00	0.69	0.08	0.36	0.51	0.56	0.28	0.75	0.65	0.73	0.31	0.50	0.62	0.58	0.47
G7	0.83	0.89	0.69	0.88	1	0.93	0.97	0.80	0.82	0.96	0.90	0.85	0.94	0.93	0.84	0.98	1	0.93	0.94	0.91	0.96
G7	57	57	54	60	57	51	53	54	56	56	57	57	55	55	52	54	57	51	57	57	57
P. parentale	0.38	0.56	0.37	0.78	0.27	0.53	0.16	0.69	0.04	0.38	0.30	0.67	0.36	0.77	0.60	0.57	0.24	0.61	0.63	0.74	0.52
G12	0.91	1	0.90	0.89	0.87	0.96	0.88	1	1	0.97	0.83	0.85	1	0.93	0.91	0.87	1	0.97	0.97	0.91	0.90
P. parentale	55	56	58	57	55	56	57	51	50	53	51	51	50	50	54	57	58	59	55	54	60
P. initiale	0.04	0.53	0.19	0.73	0.34	0.51	0.08	0.58	0.06	0.57	0.21	0.60	0.50	0.67	0.40	0.52	0.29	0.57	0.47	0.61	0.37
Le Moulon	1	1	1	0.98	1	0.97	0.79	1	1	0.82	0.94	1	0.98	1	1	1	1	0.97	1	1	1
Dhaptotypes multilocus	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

N: Nombre d'individus échantonnés par locus et par population

### III-3.2 Différenciation spatiale

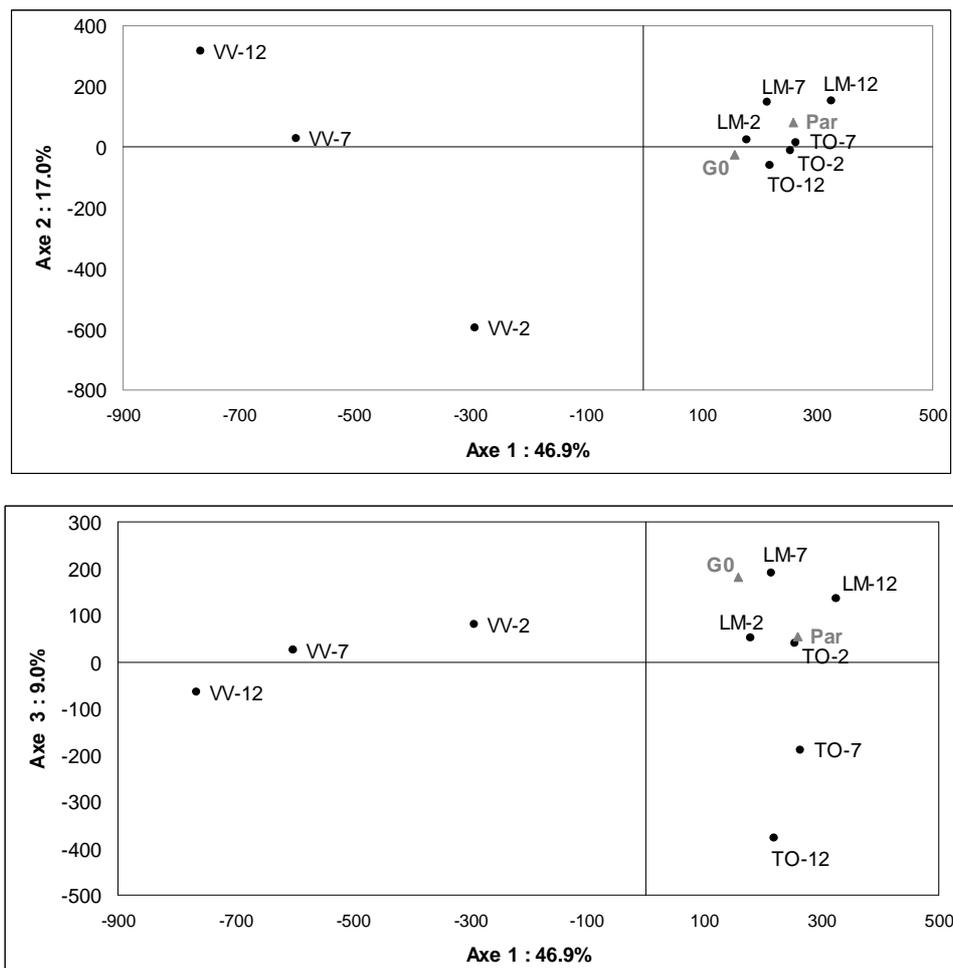


Figure III-1 : Résultats de l'analyse factorielle des correspondances (AFC). Cette analyse projette les individus sur des axes définis de façon à maximiser la variabilité inter individus, le positionnement des populations dans le plan correspond au barycentre du positionnement des individus d'une même population.

Une vision générale de la proximité génétique entre les populations est donnée Figure III-1. Les projections selon les trois premières composantes canoniques, qui expliquent en tout 72,9% de la variance totale, donnent une image de la structuration de la diversité dans l'espace (entre populations à une même génération) et dans le temps (pour une même population à différentes générations). Le premier axe sépare très nettement les populations issues de Vervins des autres populations. Cet axe ainsi que le deuxième éclatent la variabilité observée dans le temps pour la population de Vervins, tandis que le troisième axe sépare les trois générations étudiées pour la population de Toulouse. Quelque soit la projection et la génération étudiée, la population du Moulon reste génétiquement proche de la population

initiale. Cette représentation illustre le fait que les 3 populations évoluent en se différenciant les unes des autres au cours des générations.

Ces résultats issus d'analyses descriptives sont confirmés dans l'étude de la différenciation spatio-temporelle des populations. La différenciation spatiale estimée entre les trois populations à chaque génération est significative dès la première génération étudiée ( $\theta_{G2} = 0,082$ ,  $IC_{95\%}=[0,050-0,117]$ ,  $P=0,0002$ ), soit 2 générations après l'installation des populations dans les différents sites, la population la plus différenciée étant la population de Vervins (Figure III-1). Le même niveau de différenciation est observée en génération 7 ( $\theta_{G7} = 0,076$ ,  $IC_{95\%}=[0,052-0,102]$ ,  $P=0,0002$ ). En dernière génération étudiée (G12), la différenciation entre les populations accroît fortement ( $\theta_{G12} = 0,139$ ,  $IC_{95\%}=[0,110-0,167]$ ,  $P=0,0002$ ) notamment du fait de la forte différenciation de la population de Vervins comparée aux autres populations. La différenciation temporelle estimée dans chaque population est plus faible que les valeurs de différenciation spatiale ( $\theta_{LM} = 0,016$ ,  $IC_{95\%}=[0,009-0,023]$ ;  $\theta_{TO} = 0,017$ ,  $IC_{95\%}=[0,006-0,031]$ ,  $P=0,0002$ ) excepté dans la population de Vervins ( $\theta_{VV} = 0,101$ ,  $IC_{95\%}=[0,067-0,135]$ ,  $P=0,0002$ ) qui est fortement différenciée entre la première et la dernière génération étudiées ( $\theta_{VV-G2-G12} = 0,174$ ).

### **III-3.3 Variations temporelles des fréquences alléliques et effectifs efficaces**

Les variations temporelles moyennes de fréquences alléliques estimées à divers intervalles de temps à partir des variations des 21 locus microsatellites ainsi que les effectifs efficaces associés sont présentés Tableau III-3.

Pour les populations de Toulouse et Le Moulon, les effectifs efficaces estimés varient de façon relativement importante en fonction de l'intervalle de temps considéré. Notamment, les  $Ne$  estimés entre la population initiale et la génération 2 sont parmi les plus faibles, alors que les  $Ne$  estimés entre la génération 2 et la génération 7 sont les plus forts quelle que soit la population considérée. Les variations temporelles de fréquences alléliques sont du même ordre de grandeur entre ces 2 populations. Les variations temporelles de fréquences alléliques observées dans la population de Vervins sont beaucoup plus fortes que celles observées dans les populations de Toulouse et Le Moulon quelle que soit l'échelle de temps considérée, conduisant à de faibles effectifs efficaces dans cette population. Le  $Ne$  estimé entre la

population initiale et la G2 de la population de Vervins est particulièrement faible tandis que les  $N_e$  aux générations suivantes tendent à augmenter au cours du temps.

Tableau III-3 : Valeurs des variances de fréquences alléliques moyennes ( $F_c$  moyen) et effectifs efficaces moyens ( $N_e$  moyen) estimés entre 2 générations dans les différentes populations étudiées.  $t$  : nombre de générations séparant les 2 générations considérées.

	$F_c$ moyen	t	S0	St	$N_e$ moyen	IC (95%)
Par - G0	0.071	7		34	86	(46-146)
G0 - LM2	0.059	2	34	54	92	(26-2650)
LM2 - LM7	0.043	5	54	55	399	(96-2650)
LM7 - LM12	0.064	5	55	58	88	(43-193)
LM-2 - LM12	0.062	10	54	58	190	(91-439)
Par - LM12	0.080	19		58	150	(93-219)
G0 - TO2	0.069	2	34	57	45	(19-168)
TO2 - TO7	0.042	5	57	57	373	(96-2650)
TO7 - TO12	0.047	5	57	58	203	(67-2650)
TO-2 - TO12	0.079	10	57	58	113	(61-201)
Par - TO12	0.096	19		58	121	(75-175)
G0 - VV2	0.157	2	34	58	9	(6-14)
VV2 - VV7	0.117	5	58	55	31	(18-46)
VV7 - VV12	0.075	5	55	55	65	(33-125)
VV-2 - VV12	0.200	10	58	55	30	(19-43)
Par - VV12	0.253	19		55	40	(27-54)

### III-3.4 Régime de reproduction

Les valeurs de  $F_{IS}$  estimées dans chaque population sont très élevées dans l'ensemble des populations étudiées (Tableaux III-1 et III-3), comme attendu dans le cas de populations fortement autogames. Ces valeurs tendent à augmenter au cours des générations dans chacune des populations conduisant à une estimation du taux d'allogénération  $t$  de plus en plus faible dans les populations de Toulouse et Vervins en particulier. Les taux d'allogénération estimés en se basant sur le déséquilibre d'identité multilocus sont tous plus forts que les valeurs estimées à partir du  $F_{IS}$  mais ils tendent à évoluer de la même manière avec des valeurs plus faibles pour la dernière génération considérée (Tableau III-4).

Tableau III-4 : Taux d'allofécondation ( $t$ ) estimés dans les différentes populations aux différentes générations selon 2 approches : (a) approche se basant sur l'estimation de l'indice de fixation  $F_{IS}$  ; (b) approche se basant sur l'estimation du déséquilibre d'hétérozygotie multilocus ( $g_2$ , David *et al.* 2007).

Population	Le Moulon			Toulouse			Vervins		
Génération	G2	G7	G12	G2	G7	G12	G2	G7	G12
(a) $F_{IS}$	0.924	0.959	0.959	0.954	0.969	0.992	0.909	0.930	0.980
$t(F_{IS})$	<b>0.040</b>	<b>0.021</b>	<b>0.021</b>	<b>0.024</b>	<b>0.016</b>	<b>0.004</b>	<b>0.048</b>	<b>0.036</b>	<b>0.010</b>
(b) $g_2$		2.897	6.219		10.377	6.365		4.048	3.508
$s(g_2)$	NE	0.900	0.950	NE	0.969	0.951	NE	0.926	0.915
s.d.		0.037	0.020		0.016	0.051		0.024	0.058
$t(g_2)$		<b>0.100</b>	<b>0.050</b>		<b>0.031</b>	<b>0.049</b>		<b>0.074</b>	<b>0.085</b>

NE = Non Estimé

### III-3.5 Evolution du DL et génotype multilocus

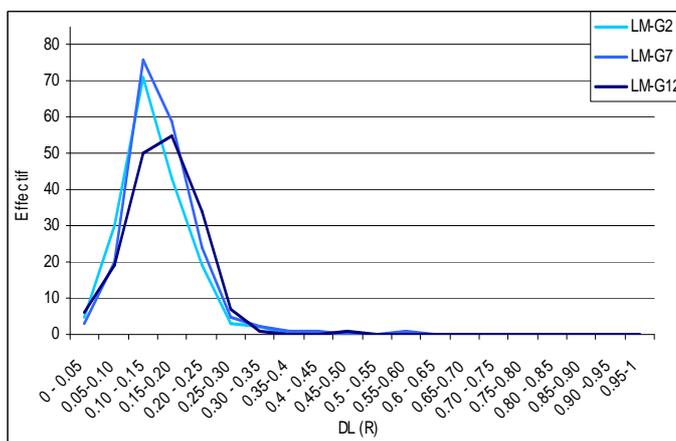
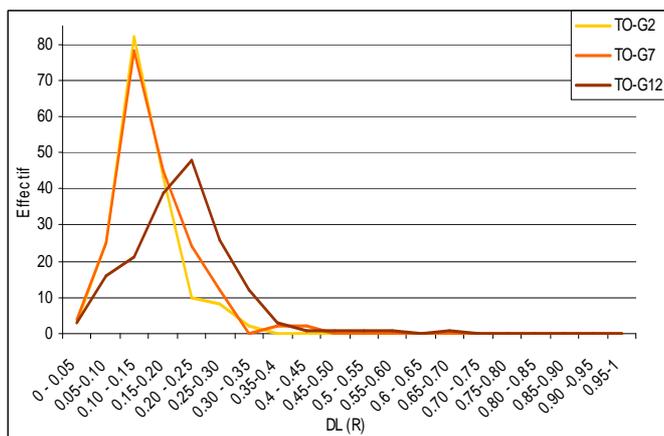
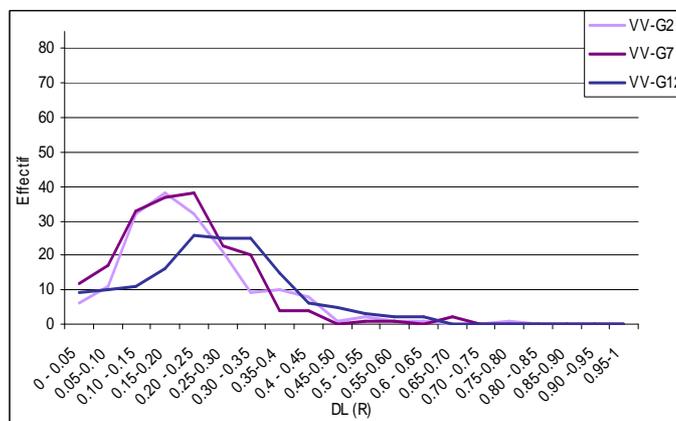
- 21 locus microsatellites

Bien que les 21 locus microsatellites étudiés soient tous localisés sur des chromosomes différents, le déséquilibre de liaison moyen estimé par la mesure de corrélation entre locus tend à augmenter au cours des générations (Tableau III-5), en particulier dans les populations de Toulouse et Vervins entre la génération 7 et la génération 12 (Figure III-2). A la génération 12, 18% des corrélations entre paires de locus polymorphes (ce qui représente 171 paires de locus testées) sont trouvées significatives au seuil de 1% dans la population du Moulon alors que cette proportion est de 39% et 32% dans les populations de Toulouse (173 comparaisons) et Vervins (155 comparaisons) respectivement.

Tableau III-5 : Niveau de corrélation moyen entre paires de locus dans les différentes populations étudiées estimé à partir des 21 locus microsatellites.

Population	Génération	R moyen	Ecart type
P. initiale	G0	0.19	0.08
Le Moulon	G2	0.14	0.05
	G7	0.16	0.06
	G12	0.16	0.06
Toulouse	G2	0.14	0.05
	G7	0.16	0.06
	G12	0.21	0.09
Vervins	G2	0.22	0.12
	G7	0.21	0.11
	G12	0.26	0.13

Figure III-2 : Distribution des valeurs de corrélations entre paires de locus polymorphes dans les 3 populations étudiées au cours des générations.



Pour comprendre comment se mettent en place ces forts déséquilibres de liaison observés entre locus, des génotypes multilocus ont été reconstitués à partir des 14 locus présentant le moins de données manquantes (Tableau III-1, Figure III-3).

400 haplotypes ont été trouvés sur les 397 génotypes complets du jeu de données exploitables pour la reconstruction des haplotypes. Aucun des 16 haplotypes parentaux n'a été retrouvé dans les populations échantillonnées. Seuls 2 des 28 haplotypes de la population initiale ont été retrouvés dans une population : celle de Toulouse. Seuls 4 haplotypes sont communs à des populations issues de sites différents (3 communs à Toulouse et Le Moulon, 1 commun à Toulouse et Vervins). Toutes générations confondues, 119, 123 et 110 haplotypes sont spécifiques des populations du Moulon, de Toulouse et de Vervins respectivement. Intra population, la majorité des haplotypes observés en G12 ne sont pas observés avant dans la population : dans la population du Moulon, 39 des 43 haplotypes observés en génération 12

ne sont pas observés aux générations précédentes, à Toulouse et Vervins respectivement 28 sur 37 et 22 sur 27 haplotypes observés en G12 sont des haplotypes nouveaux comparés à ceux trouvés aux générations précédentes.

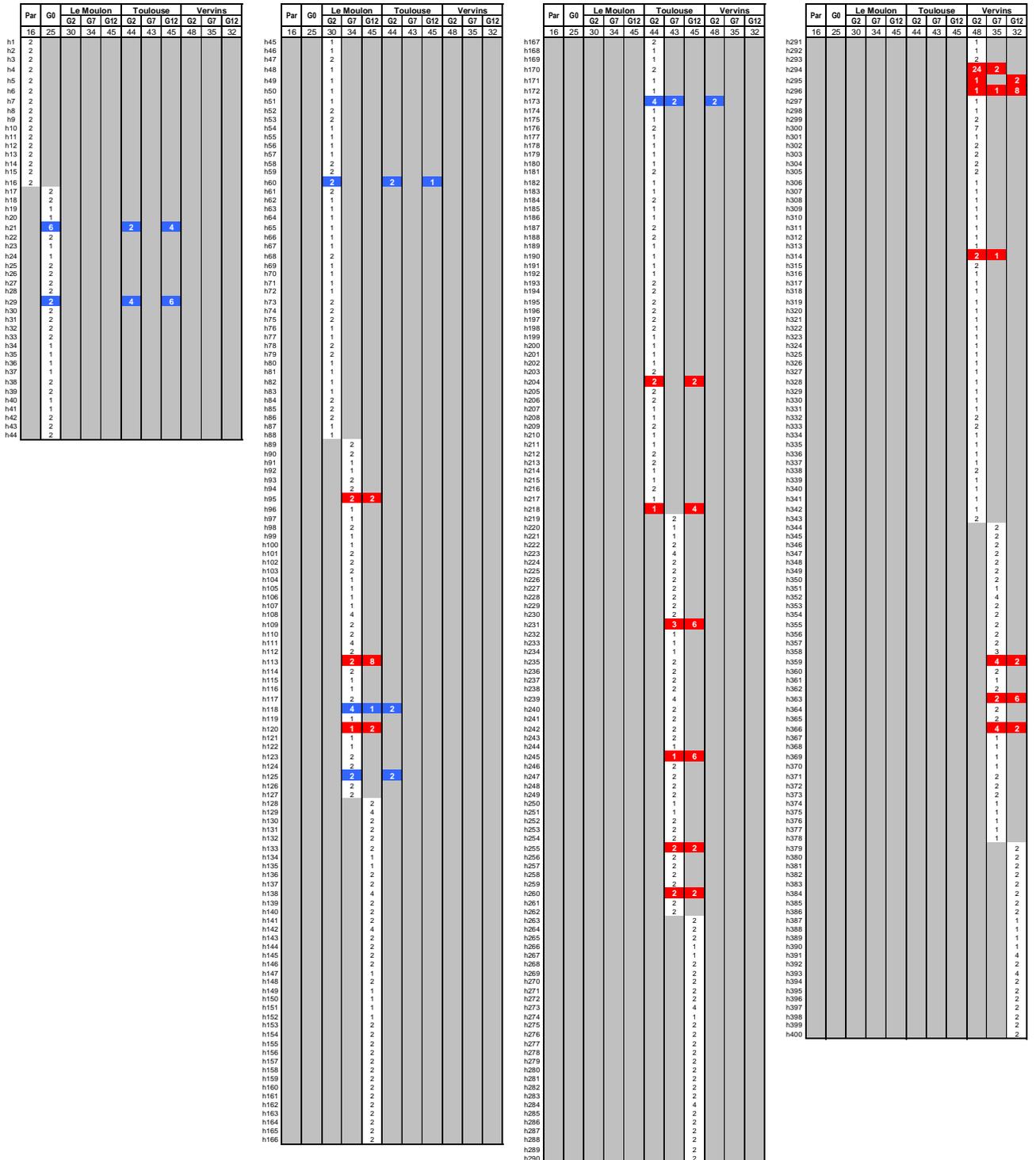


Figure III-3 : Ensemble des haplotypes multilocus reconstruits à partir de 14 locus microsatellites non liés et répartition dans les populations étudiées. En rouge sont indiqués les haplotypes présents dans une seule population échantillonnée à des générations différentes, et en bleu les haplotypes présents dans 2 populations issues de 2 sites différents.

Tableau III-6 : Nombre de génotypes complets utilisés pour la reconstruction d'haplotypes et nombre d'haplotypes trouvés dans les trois populations au cours des générations.

Population	Le Moulon			Toulouse			Vervins		
Génération	G2	G7	G12	G2	G7	G12	G2	G7	G12
Nombre de génotypes complets utilisés pour la reconstruction d'haplotypes	30	34	45	44	43	45	48	35	32
Nombre d'haplotypes trouvés dans la population	44	39	43	57	45	37	54	38	27
Nombre d'haplotypes rapporté au nombre de génotypes considérés	1.47	1.15	0.96	1.30	1.05	0.82	1.13	1.09	0.84

Pour l'ensemble des populations, le nombre d'haplotypes rapporté au nombre de génotypes utilisés dans cette analyse tend à diminuer au cours des générations (Tableau III-6). Dans les premières générations, la présence d'hétérozygotes à au moins un locus augmente le nombre d'haplotypes trouvés tandis que dans la dernière génération étudiée, certains haplotypes sont retrouvés plusieurs fois à l'état homozygote. Dans la population de Vervins, 12 individus sur les 48 utilisés pour la reconstruction d'haplotypes présentent les mêmes associations alléliques à tous les locus en génération 2 (h294, Figure III-3). Un autre haplotype est observé 7 fois (3 génotypes homozygotes et un hétérozygote, h300) dans la population de Vervins en génération 2. Ces 2 haplotypes majoritaires dans la population de Vervins en génération 2 présentent la même composition allélique à 13 des 14 locus considérés. L'ensemble des allèles constituant ces 2 haplotypes majoritaires sont des allèles initialement observés dans la population parentale. Ces 2 haplotypes majoritaires ne sont pas retrouvés en dernière génération étudiée tandis que d'autres haplotypes deviennent plus fréquents. Ceci témoigne d'importants remaniements au sein de cette population au cours des générations.

Afin de mieux appréhender les événements de recombinaison qui ont eu lieu et l'étendue du déséquilibre de liaison le long d'un chromosome dans les 3 populations étudiées, une étude plus fine à l'échelle d'un chromosome a ensuite été entreprise à l'aide de 7 marqueurs répartis le long du chromosome 3B.

#### - 7 locus microsatellites du chromosome 3B

Les 7 locus microsatellites présentent un niveau de diversité comparable à la diversité estimée sur les 21 locus microsatellites répartis sur l'ensemble du génome. Le nombre moyen d'allèles par locus, toutes populations confondues, est de 3,80. Comme pour les 21 locus microsatellites, de nouveaux allèles apparaissent au cours des générations dans les différentes populations.

Tableau III-7 : Estimation par locus des paramètres de diversité des 7 locus microsatellites positionnés sur le chromosome 3B : nombre d'allèles et diversité de Nei ( $H_e$ ) par population.

Locus			<i>Xgym566</i>	<i>Xwmc231</i>	<i>Xwmc777</i>	<i>Xbarc164</i>	<i>Cfi53</i>	<i>Xwmc291</i>	<i>Cfi3366</i>	
Nombre d'allèles parentaux			5	4	3	4	2	4	2	
Nombre total de nouveaux allèles			1	2	0	4	0	4	0	
Nombre total d'allèles en	Le Moulon	G12	6	4	3	4	2	6	2	
	Toulouse	G12	6	3	3	6	2	3	2	
	Vervins	G12	4	4	2	5	2	7	2	
Paramètres estimés par population: N $H_e$	P. parentale		16	16	16	16	16	16	16	
			<b>0.75</b>	<b>0.58</b>	<b>0.65</b>	<b>0.67</b>	<b>0.43</b>	<b>0.63</b>	<b>0.43</b>	
	P. initiale G0		31	34	36	33	36	32	33	
			<b>0.64</b>	<b>0.56</b>	<b>0.66</b>	<b>0.71</b>	<b>0.48</b>	<b>0.37</b>	<b>0.45</b>	
	Le Moulon	G2	56	58	58	57	57	52	56	
				<b>0.69</b>	<b>0.53</b>	<b>0.62</b>	<b>0.63</b>	<b>0.50</b>	<b>0.62</b>	<b>0.45</b>
		G7	51	55	54	58	57	54	60	
			<b>0.74</b>	<b>0.63</b>	<b>0.67</b>	<b>0.74</b>	<b>0.50</b>	<b>0.67</b>	<b>0.45</b>	
		G12	57	59	59	59	60	57	58	
			<b>0.67</b>	<b>0.65</b>	<b>0.67</b>	<b>0.71</b>	<b>0.44</b>	<b>0.64</b>	<b>0.49</b>	
	Toulouse	G2	55	51	58	60	60	59	60	
				<b>0.75</b>	<b>0.55</b>	<b>0.67</b>	<b>0.63</b>	<b>0.50</b>	<b>0.53</b>	<b>0.44</b>
		G7	58	57	57	58	59	57	55	
			<b>0.75</b>	<b>0.51</b>	<b>0.67</b>	<b>0.62</b>	<b>0.50</b>	<b>0.56</b>	<b>0.41</b>	
		G12	56	57	59	56	59	58	56	
		<b>0.62</b>	<b>0.34</b>	<b>0.55</b>	<b>0.47</b>	<b>0.47</b>	<b>0.32</b>	<b>0.38</b>		
Vervins	G2	56	59	59	60	57	58	59		
			<b>0.69</b>	<b>0.16</b>	<b>0.50</b>	<b>0.58</b>	<b>0.46</b>	<b>0.64</b>	<b>0.18</b>	
	G7	53	55	54	55	57	56	52		
		<b>0.69</b>	<b>0.20</b>	<b>0.44</b>	<b>0.67</b>	<b>0.36</b>	<b>0.59</b>	<b>0.21</b>		
	G12	51	51	57	58	58	59	58		
		<b>0.58</b>	<b>0.25</b>	<b>0.31</b>	<b>0.68</b>	<b>0.40</b>	<b>0.54</b>	<b>0.13</b>		

N: Nombre d'individus échantillonnés par locus et par population

Les valeurs moyennes de déséquilibre de liaison sur les 7 locus du chromosome 3B estimées dans les différentes populations (Figure III-8) sont du même ordre que les valeurs estimées sur les 21 locus répartis sur l'ensemble du génome (Tableau III-5). Comme l'illustre la Figure III-4, une part de ces déséquilibres de liaison est clairement liée à la distance séparant les locus dans les populations du Moulon et de Toulouse : plus la distance entre 2 locus est faible, plus le déséquilibre de liaison est important (corrélation entre la distance génétique séparant 2 locus et leur niveau de déséquilibre de liaison de -0,51 et -0,63 pour les populations du Moulon et de Toulouse respectivement en génération 12 pour les locus distants de moins de 30 cM). Cette corrélation est moins claire dans la population de Vervins (corrélation de -0,22). Les déséquilibres de liaison entre locus très distants sont quant à eux du même ordre que la valeur moyenne estimée à partir des 21 microsatellites répartis sur l'ensemble du génome.

Tableau III-8 : Valeurs moyennes de corrélation entre paires de locus dans les différentes populations étudiées estimées à partir des 7 locus microsatellites du chromosome 3B.

Population	Génération	<i>R</i> moyen	Ecart type
P. initiale	G0	0.19	0.10
Le Moulon	G2	0.14	0.06
	G7	0.16	0.06
	G12	0.20	0.07
Toulouse	G2	0.15	0.06
	G7	0.14	0.05
	G12	0.19	0.11
Vervins	G2	0.17	0.10
	G7	0.21	0.13
	G12	0.25	0.16

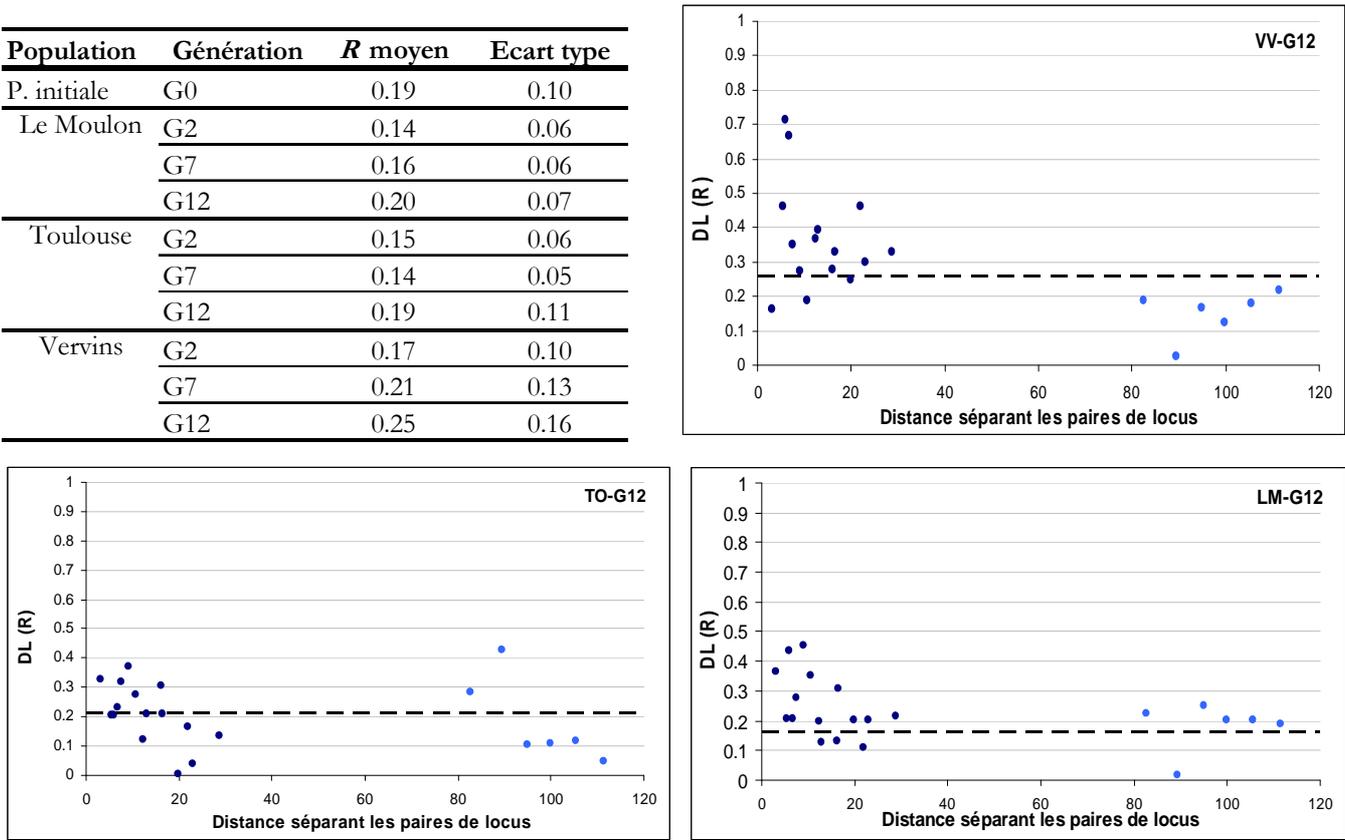


Figure III-4 : Déséquilibre de liaison moyen estimé entre paires de locus en fonction de la distance génétique séparant ces paires de locus pour les 3 populations en génération 12. La ligne en pointillés gris représente la valeur moyenne du DL estimé sur les 21 locus répartis sur l'ensemble du génome.

La reconstruction d'haplotypes multilocus conduit à l'observation de 239 haplotypes sur les 442 individus considérés dans cette analyse. Du fait de la grande diversité initiale observée aux locus microsatellites, 12 haplotypes sont observés parmi les 16 génotypes parentaux (Figure III-5). De nouveaux haplotypes apparaissent dans la population initiale G0 par recombinaison (18 haplotypes sur les 23 observés) et du fait de l'apparition de nouveaux allèles à certains locus (2 haplotypes). Seuls 3 haplotypes parentaux sont retrouvés dans cette population.

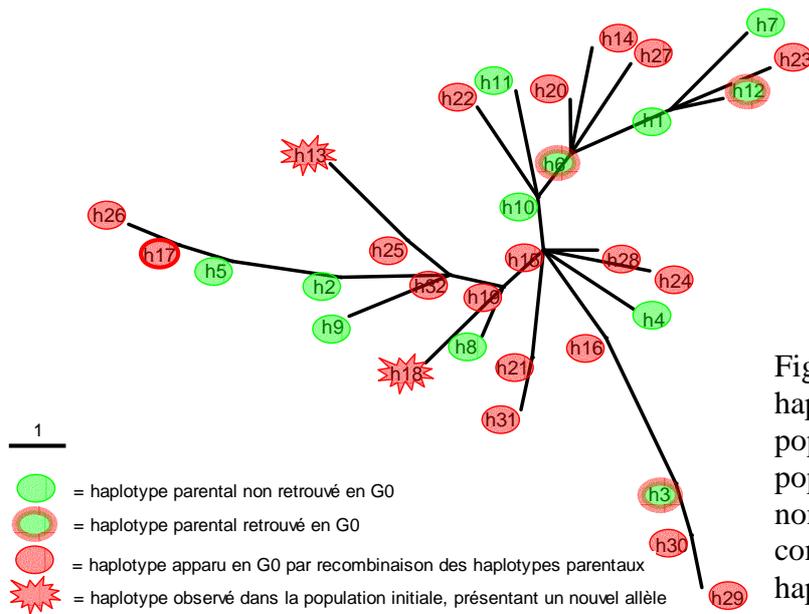
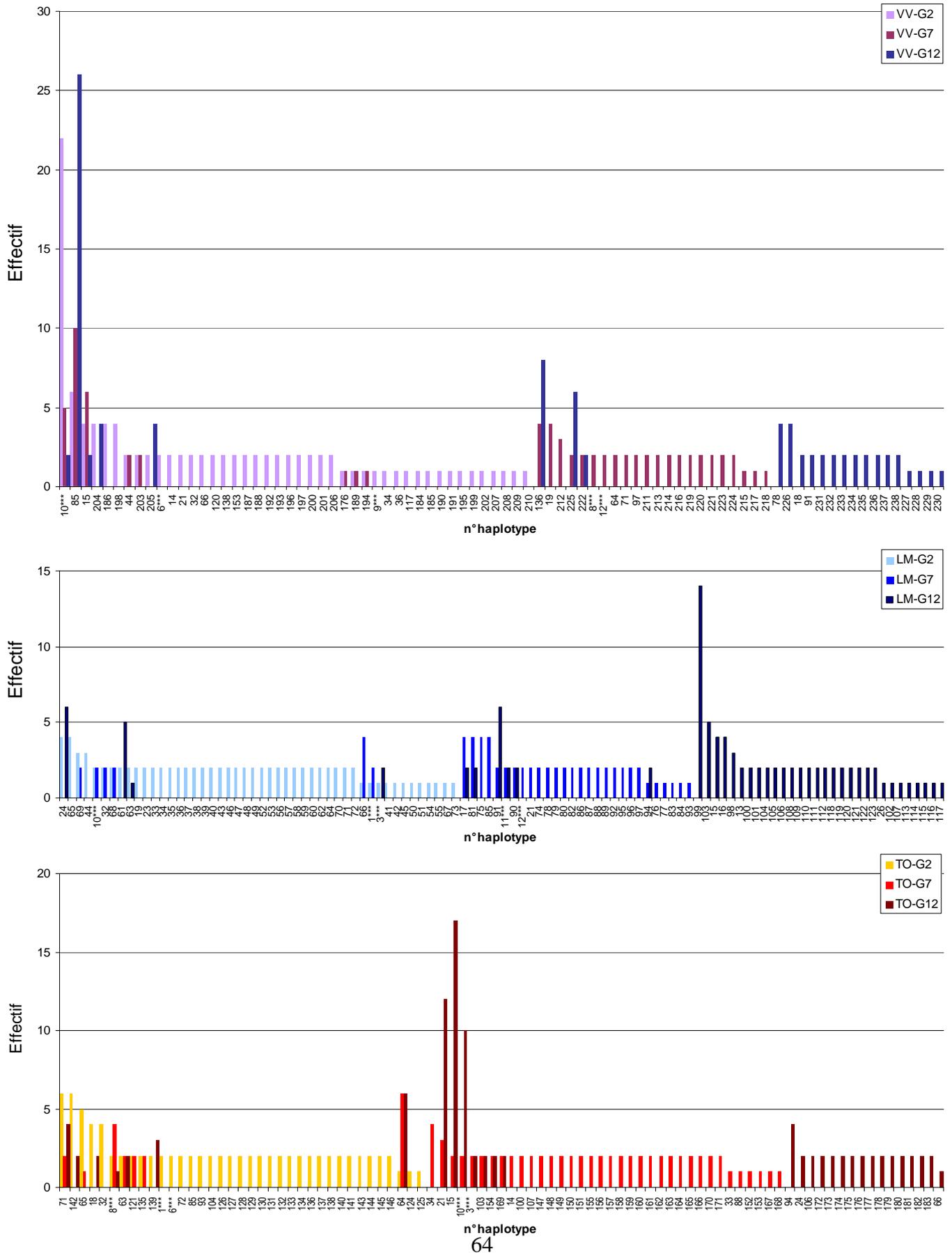


Figure III-5 : Réseau des haplotypes trouvés dans la population parentale et dans la population initiale basé sur le nombre de différences dans la composition allélique des haplotypes.

132 haplotypes sont retrouvés dans la population du Moulon toutes générations confondues, parmi lesquels 5 sont des haplotypes parentaux, 12 sont des haplotypes de la population initiale et 69 sont des haplotypes trouvés exclusivement dans cette population. Parmi les 137 haplotypes trouvés dans la population de Toulouse (131 dans la population de Vervins), 5 (5, VV) sont des haplotypes parentaux, 8 (8, VV) sont des haplotypes de la population initiale et 56 (56, VV) des haplotypes trouvés exclusivement dans cette population.

L'évolution des fréquences haplotypiques représentée Figure III-6 (voir aussi annexe 2) montre que certains haplotypes augmentent en fréquences de façon importante dans les différentes populations. Ainsi dans les populations du Moulon et de Toulouse, respectivement 1 et 3 haplotypes sont retrouvés majoritaires en génération 12. Pour Toulouse, ces haplotypes majoritaires étaient déjà présents à la génération 7. Dans la population de Vervins, un haplotype déjà observé dans la population parentale, est trouvé à une forte fréquence (11 génotypes homozygotes, génotypes communs aux 12 trouvés avec le même haplotype pour les 21 locus) dès la génération 2, mais cet haplotype diminue très fortement en fréquence aux cours des générations, au profit d'un haplotype déjà observé à la première génération étudiée. Ainsi des blocs haplotypiques le long du chromosome 3B apparaissent et augmentent fortement dans chacune des populations.

Figure III-6 : Evolution de la fréquence des haplotypes multilocus pour les 7 locus microsatellites du chromosome 3B dans les 3 populations observées au cours des générations. \*\*\* indique que cet haplotype était déjà observé dans la population parentale (qui comptabilise au totale 12 haplotypes différents).



## III-4. Discussion

### III-4.1 Régime de reproduction

Le régime de reproduction peut grandement influencer l'évolution et la structuration de la diversité neutre dans les populations. C'est pourquoi il est nécessaire de préciser la situation des populations étudiées sur ce point en premier lieu.

Comme attendu dans le cas des populations expérimentales de blé, les valeurs du coefficient de consanguinité,  $F_{IS}$ , sont très élevées pour l'ensemble des populations étudiées, suggérant un déficit en hétérozygotes important, lié probablement à un régime de reproduction majoritairement autogame. De fait, les taux d'allo-fécondation estimés à partir du  $F_{IS}$  sont faibles pour toutes les populations en particulier dans la dernière génération considérée (Tableau III-3).

Les taux d'allofécondation estimés en G2 sont plus élevés que les valeurs trouvées en G7 et G12 pour l'ensemble des populations. Cependant cette valeur estimée est discutable puisque les plus faibles valeurs de  $F_{IS}$  trouvées en génération 2 peuvent être liées à la fois (i) à de l'hétérozygotie résiduelle générée par les croisements forcés effectués entre les génotypes parentaux pour constituer la population initiale (4 générations de croisements forcés suivies de 3 générations de fécondation libre), et (ii) à de l'hétérozygotie générée par des événements d'allofécondation en G1. En revanche, du fait du mode de reproduction majoritairement autogame, on peut estimer que cette confusion est négligeable en génération 7 et en génération 12 puisque respectivement 10 et 15 générations séparent ces générations du dernier croisement forcé réalisé pour constituer la population initiale.

Les taux d'allofécondation estimés dans les populations du Moulon et de Toulouse en G12 ( $t_{LM-G12} = 2,1\%$  ;  $t_{TO-G12} = 0,4\%$ ) tendent à être plus faibles que les valeurs estimées en G10 selon la même méthode et dans les mêmes populations à partir de marqueurs RFLP ( $t_{LM-G10} = 5,9\%$  ;  $t_{TO-G10} = 10,1\%$  ; Enjalbert *et al.* 1998). Cette différence peut être attribuée au type de marqueurs utilisé plus ou moins sujet aux erreurs de génotypage. En effet, l'utilisation du  $F_{IS}$  pour estimer le taux d'allofécondation suppose que le taux d'hétérozygotes est bien estimé dans les populations. Or un nombre croissant d'études suggère que les erreurs de génotypage sont loin d'être négligeables et peuvent avoir un impact important sur certains estimateurs

classiquement utilisés en génétique des populations (Hoffman et Amos 2005, Pompanon *et al.* 2005, Bonin *et al.* 2004). Notamment, le taux d'hétérozygotes pourrait être sous-estimé au cours du génotypage du fait de mécanismes désormais bien connus tels que la présence d'allèles nuls ou la compétition d'amplification d'un des 2 allèles présents chez un individu hétérozygote lors de la PCR, conduisant à l'amplification exclusive du fragment le plus court (phénomène dit de dominance partielle, Wattier *et al.* 1998). Lors de la lecture des données de génotypage microsatellite issues de la migration sur gel d'acrylamide des fragments PCR, l'identification des hétérozygotes a été particulièrement difficile. D'une part, les allèles à l'état hétérozygote tendent à produire un signal de plus faible intensité qu'à l'état homozygote, conduisant dans certain cas à des niveaux d'intensité situés en dessous du seuil de détection : de ce fait, sur l'ensemble des locus, une plus grande part d'hétérozygotes est susceptible d'être lue comme donnée manquante par rapport aux homozygotes. D'autre part, lors de l'amplification de locus microsatellites, des sous-produits d'amplification peuvent apparaître au cours de la PCR conduisant à des profils multi-pics (stutter) rendant la détection d'hétérozygotes complexe dans certains cas (Figure III-7). Il est donc probable que le taux d'allofécondation estimé à partir des marqueurs microsatellites soit sous-estimé du fait de la sous-estimation du taux d'hétérozygotes.

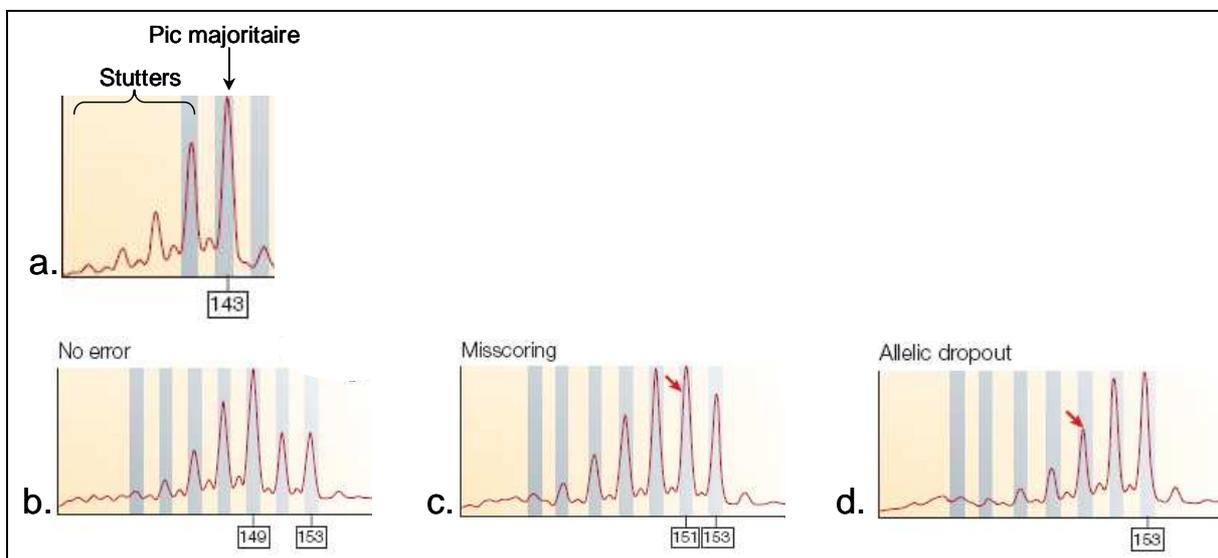


Figure III-7 : Exemple de profils de fluorescence observables suite à la migration sur séquenceur capillaire de produits d'amplification d'un locus microsatellite. *a.* Profil microsatellite « type », présentant un pic majoritaire suivi de sous-produits d'amplification de plus faible intensité (stutter) lié au dérapage de la polymérase ; *b-d.* Différents profils observables pour le même génotype hétérozygote : *b.* les 2 allèles se distinguent bien l'un de l'autre ; *c.* la présence de stutters de forte intensité conduit à une mauvaise identification d'un allèle chez l'hétérozygote ; *d.* l'allèle de plus petite taille, de plus faible intensité, n'est pas détecté du fait de la confusion avec un stutter (d'après Pompanon *et al.* 2005).

C'est pourquoi nous avons également utilisé une méthode développée récemment (David *et al.* 2007) qui s'affranchit de ces biais de génotypage en se basant sur la distribution multilocus des hétérozygotes pour estimer le taux d'autofécondation. Notons que l'utilisation de cette méthode, comme celle de la méthode par estimation du  $F_{IS}$ , suppose que le régime de reproduction est stable et à l'équilibre de consanguinité. Cette méthode ne nous permet donc pas plus que la précédente d'estimer de façon fiable la nature du régime de reproduction à la première génération étudiée. De plus, dans le cas des populations de blé soumises à la stochasticité du milieu, la stabilité du régime de reproduction ne peut être assurée car il est connu chez le blé que les facteurs environnementaux peuvent avoir un impact sur le régime de reproduction. En particulier, une faible luminosité associée à une faible température lors de la méiose peut entraîner des problèmes de stérilité mâle augmentant ainsi le taux d'allofécondation (Demotes-Mainard *et al.* 1995). Elle suppose de plus que tous les locus sont affectés par des erreurs de génotypage de la même façon. Cependant du fait de l'utilisation de marqueurs microsatellites, il semble que les locus à l'état hétérozygote soient particulièrement affectés par des erreurs de lecture (*cf.* arguments présentés précédemment) ayant pu conduire à la sous-estimation des locus hétérozygotes ou à leur classement majoritaire en données manquantes. Dans ce cas, les valeurs d'alfofécondation estimées par cette méthode peuvent être sous-estimées. Enfin, les auteurs mettent en garde quant à son usage lorsque les locus utilisés sont en déséquilibre de liaison. En particulier, dans le cas d'une population fortement autogame, présentant un faible effectif efficace ( $N_e < 100$ ) et un faible taux de recombinaison efficace, de forts déséquilibres de liaison entre locus peuvent apparaître, pouvant conduire à une sous-estimation du taux d'alfofécondation (David *et al.* 2007). Ce risque est particulièrement important dans le cas de la population de Vervins qui, en plus d'avoir un effectif efficace faible, présente d'importants déséquilibres de liaisons entre locus. Cependant, le taux d'alfofécondation estimé sur la base du déséquilibre d'identité dans cette population est le plus fort parmi toutes les populations étudiées en G12, ce qui laisse supposer que le taux d'alfofécondation n'est pas sous estimé.

Toutefois, malgré ces difficultés rencontrées pour estimer le taux d'alfofécondation dans les différentes populations et son évolution au cours des générations, il semble clair que les populations étudiées présentent toutes un régime de reproduction majoritairement autogame accompagnée d'une part d'alfofécondation résiduelle de l'ordre de 5%.

### III-4.2 Régime de reproduction et effectif efficace

Le régime de reproduction décrit précédemment a une conséquence directe sur l'effectif efficace attendu : dans une population majoritairement autogame, les individus sont principalement homozygotes et le nombre de gamètes indépendants disponibles pour la reproduction est réduit de moitié comparé à une population majoritairement allogame (Glémin *et al.* 2006). Un effectif efficace peut être calculé sur la base de l'effectif démographique observé à chaque génération, corrigé par le coefficient de consanguinité ( $N_{ed} = N / (1 + F_{is})$ ). Dans les populations de gestion dynamique, au minimum 5000 graines sont semées chaque année pour constituer une génération dont l'effectif efficace démographique estimé de cette façon est légèrement supérieur à 2500. Les effectifs efficaces estimés sur la base des variations de fréquences alléliques des 21 marqueurs microsatellites entre différents intervalles de temps sont pourtant bien plus faibles pour l'ensemble des populations étudiées (compris entre 40 et 150 pour un  $N_e$  estimé entre la population parentale et la génération 12). Pour les populations du Moulon et de Toulouse, les valeurs de  $N_e$  observées dans cette étude sont proches de celles précédemment estimées dans les populations de GD à partir de marqueurs RFLP (de l'ordre de 150, Goldringer *et al.* 2001, Encadré III-1). Dans une méta-analyse incluant 192 observations d'effectifs efficaces dans des populations naturelles de divers organismes, Frankham (1995) montre que le ratio  $N_e/N$  est proche de 0,10. Siol *et al.* (2007) a récemment montré que ce ratio était compris entre 0,5 et 0,10 chez une espèce très fortement autogame, *Medicago truncatula*. Le rapport « effectif efficace / effectif démographique » trouvé pour les populations du Moulon et de Toulouse est dans cette fourchette de valeurs (0,06 pour Le Moulon ; 0,05 pour Toulouse pour les  $N_e$  estimés entre la population parentale et la G12), tandis que ce rapport est de 0,01 pour la population de Vervins. La situation particulière de cette population sera discutée ultérieurement.

Les faibles effectifs efficaces observés peuvent s'expliquer par de fortes différences, héritables ou non héritables, dans la contribution de chaque individu de la population à la constitution de la génération suivante. Frankham (1995) souligne que 2 facteurs semblent affecter de façon importante l'effectif efficace d'une population : (i) des fluctuations de taille de population d'une génération à l'autre, et (ii) une variance importante du succès reproducteur entre famille. Des variations importantes de taille de population sont susceptibles d'abaisser ponctuellement l'effectif efficace d'une population en accentuant le phénomène de dérive. Dans les cas des populations de gestion dynamique, le nombre d'individus semé à chaque génération est contrôlé par le dispositif expérimental. Un accident lors du semis ou lors de la

croissance des plantes peut ponctuellement affecter le nombre d'individus participant à la production de la génération suivante mais ces variations ponctuelles non hérissables ne peuvent expliquer les faibles effectifs observés dans toutes les populations à toutes les générations. Ce faible effectif efficace semble donc principalement dû à de fortes différences de contribution entre individus pour la constitution de la génération suivante qui sont liées aux succès reproducteur de ces individus. En effet, un individu présentant une bonne valeur sélective dans un environnement donné est plus susceptible de participer à la génération suivante qu'un individu moins bien adapté. L'échantillonnage des gamètes pour constituer la génération suivante n'est alors pas le simple fait du hasard (*i.e.* de la dérive) mais implique aussi un processus de sélection des gènes d'adaptation. L'hypothèse de sélection affectant l'ensemble du génome est également étayée par le fort niveau de différenciation observé entre les populations. En effet, le niveau de différenciation estimé en génération 12 pour les 3 populations ( $\theta = 0,14$ ) est largement supérieur au niveau de différenciation attendu dans le cas de populations de même origine évoluant de façon isolée sous le seul effet de la dérive ( $\theta = 1 - e^{-t/2N} = 0,002$  avec  $t=12$  et  $N=2650$ , effectif démographique corrigé par le coefficient de consanguinité, Hedrick 2005).

Pour mieux comprendre comment ces phénomènes de sélection peuvent affecter l'évolution de l'ensemble du génome, il est nécessaire d'examiner de plus près le déséquilibre de liaison et l'évolution des associations alléliques dans ces populations.

### III-4.3 Régime de reproduction et déséquilibre de liaison

Au cours des générations, de forts déséquilibres de liaison tendent à se mettre en place tant entre locus situés sur un même chromosome qu'entre les locus microsatellites situés sur des chromosomes différents, dans les 3 populations étudiées. Cette observation est une conséquence directe du régime de reproduction et des faibles effectifs efficaces des populations. En effet, un fort niveau d'homozygotie est attendu à chaque locus dans le cas de populations autogames ce qui réduit le nombre d'évènements de recombinaison efficaces qui rompraient la rupture d'association entre 2 locus.

A l'échelle d'un chromosome, de forts déséquilibres de liaison ont été observés entre locus proches (Figure III-4). Ce déséquilibre tend à décroître avec la distance génétique jusqu'à une distance de 10 à 20 cM. La reconstruction d'haplotypes montre que des blocs haplotypiques sont créés au cours des générations le long de tout le chromosome, conduisant ainsi à la

formation d'un déséquilibre à longue distance. La décroissance du déséquilibre de liaison en fonction de la distance entre locus est un phénomène très étudié chez divers organismes car l'étendue du DL a un impact direct sur la résolution des études qui visent à identifier les locus impliqués dans la variation phénotypique d'un caractère (Rafalski 2002). L'étendue du déséquilibre de liaison observée dans cette étude est du même ordre de grandeur que les valeurs estimées chez des plantes autogames cultivées (Chao *et al.* 2007 : DL significatif jusqu'à une distance de 20 cM dans des variétés américaines de blé hexaploïde ; Kraakman *et al.* 2004 : étendue du DL observée jusqu'à 10 cM pour des variétés d'orge).

A l'échelle du génome, certaines combinaisons alléliques sont retrouvées chez plusieurs individus à la dernière génération observée (Figure III-3), ce qui suggère qu'une structuration en lignées s'est mise en place à l'intérieur même des populations. L'apparition de ces lignées est probablement due à des individus initialement très hétérozygotes qui, du fait des croisements artificiels effectués pour constituer la population initiale G<sub>0</sub>, se sont reproduits de façon quasi-exclusive par autofécondation jusqu'à une fixation complète de l'ensemble des locus. Les effets combinés de la dérive, liée au faible effectif efficace, et de la sélection, peuvent également accélérer l'apparition de cette structuration en lignées. Dans un contexte autogame, on s'attend à ce que les individus combinant le plus d'allèles favorables soient rapidement sélectionnés. En effet, l'homozygotie accroît la probabilité de fixation des allèles favorables (Charlesworth 2003). Néanmoins, la sélection dans des populations autogames de taille finie tend à amplifier la dérive qui entraîne alors de forts déséquilibres de liaison entre les allèles favorables et les allèles défavorables (David *et al.* 1993). Dans ce contexte, des allèles faiblement délétères peuvent être rapidement fixés (Charlesworth 2003) et la sélection finit par perdre en efficacité à plus long terme (Glémin *et al.* 2006). Les événements d'allofécondation peuvent rompre ces associations entre allèles générant de nouvelles combinaisons alléliques potentiellement mieux adaptées.

Dans le cas de la population de gestion dynamique, de nombreuses combinaisons alléliques étaient disponibles dès la population initiale du fait du système multi-parental et du brassage important généré par les croisements forcés entre les lignées de départ. Cependant le taux d'allofécondation résiduel finalement observé dans les populations a permis de générer dans les différentes populations un grand nombre d'haplotypes qui n'étaient pas dans la population initiale, et ce sont principalement ces haplotypes qui augmentent en fréquence (Figure III-3 et Figure III-6).

### III-4.4 Nouveaux allèles : mutation ou migration ?

L'apparition de nouvelles combinaisons alléliques au niveau des marqueurs microsatellites dans ces populations s'explique également par l'apparition de nouveaux allèles à différents locus en particulier dans la population initiale comparée à la population parentale. Ces nouveaux allèles ont pu apparaître par mutation ou par migration, bien que les populations aient été cultivées dans des conditions *a priori* isolées des autres cultures de blé des différents sites. L'identification du processus réel d'émergence de ces nouveaux allèles est importante car si la mutation n'affecte pas tous les locus de la même façon, la migration, elle, est susceptible d'apporter une nouvelle diversité à l'ensemble des locus.

Les taux de mutation estimés à partir de locus microsatellites sont souvent plus forts que les taux estimés sur d'autres types de marqueurs avec des valeurs de l'ordre de  $10^{-3}$  à  $10^{-4}$  (Raquin 2005, Thuillet *et al.* 2002, Vigouroux *et al.* 2002). L'origine du fort taux de mutation de ce type de marqueurs a été bien documentée. Deux mécanismes induisant une variation du nombre de motifs constituant le microsatellite sont généralement invoqués : le dérapage de la polymérase lors de la réplication de l'ADN et des crossing-over inégaux lors de la reproduction. Une étude détaillée de l'apparition de nouveaux allèles à des locus microsatellites dans la population de gestion dynamique du Moulon après 17 générations d'évolution a été menée dans une étude précédente et favorise l'hypothèse mutationnelle (Raquin 2005, Raquin *et al.* en préparation). Plusieurs arguments sont invoqués : (i) de nouveaux allèles sont apparus au cours des générations étudiées (G1, G5, G10 et G15 de la population du Moulon), (ii) aucun des nouveaux allèles échantillonnés dans la population ne semble provenir des variétés les plus couramment cultivées sur le site, (iii) le séquençage des nouveaux allèles montrent que ces allèles diffèrent des allèles présents chez les parents par le nombre de répétitions du motif microsatellite plutôt que par les régions flanquantes.

Dans notre étude, un grand nombre de nouveaux allèles sont trouvés dans la population initiale (17 allèles) ou sont retrouvés dès la génération 2 dans au moins 2 des 3 populations étudiées (3 allèles), ce qui laisse penser que ces allèles étaient présents dans la population initiale mais n'ont pas été échantillonnés. Raquin (2005) suggère que l'apparition d'un grand nombre de nouveaux allèles aux premières générations serait liée à la forte hétérozygotie générée par les croisements forcés effectués pour constituer la population initiale. En effet, une forte hétérozygotie tend à augmenter l'instabilité des microsatellites en augmentant la probabilité de crossing-over inégaux (Rubinsztein *et al.* 1995). A la différence de l'étude menée par Raquin (2005), nous disposons dans la notre de données de génotypage sur des

individus issus de la population initiale G0. Sur les 36 individus génotypés, 12 présentent un nouvel allèle à l'un des 21 locus microsatellites, 3 cumulent 2 nouveaux allèles, et un génotype cumule 5 allèles sur l'ensemble des 21 locus. L'apparition par mutations indépendantes de 5 nouveaux allèles dans un même génotype est peu probable (la probabilité d'apparition indépendante de 5 nouveaux allèles dans G0 est égale à 0,0002, Figure III-8). Ceci laisse supposer que la population initiale a été contaminée, soit par du pollen extérieur lors de la multiplication soit par des graines lors de la récolte.

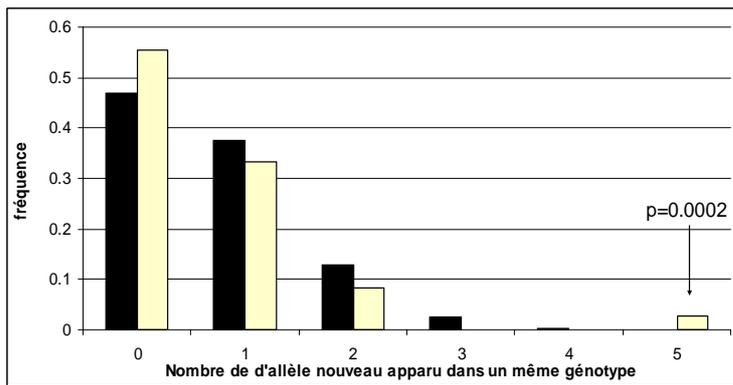


Figure III-8 : Distribution des probabilités d'apparition des nouveaux allèles de façon indépendante dans un même génotype estimé à partir de la fréquences des nouveaux allèles (en noir) comparé aux fréquences observées de génotypes cumulant 0, 1, 2, *etc.*, nouveaux allèles dans la population initiale G0 (en clair)

Cependant, l'apparition de nouveaux allèles de façon spécifique dans les différentes populations au fil des générations laissent supposer que la mutation joue également un rôle important dans l'apparition d'une nouvelle diversité aux locus microsatellites. Du fait de la culture en isolement des populations dans les différents sites (au moins 50 mètres séparent les populations d'autres cultures de blé), la contamination par du pollen extérieur est peu probable (Rhoné et *al.* 2008). Un suivi plus fin de l'évolution du polymorphisme microsatellite entre des individus de la génération 16 et leur descendance en génération 17 dans la population du Moulon montre que le taux de mutation estimé par observation directe serait compris entre 0 et  $6.10^{-2}$  selon les locus (Raquin 2005). Il semble donc finalement que les nouveaux allèles observés dans les différentes populations de notre étude aient deux origines distinctes : la mutation induisant l'apparition de nouveaux allèles au fil des générations et la contamination de la population initiale.

### III-4.5 Cas particulier de la population de Vervins

Nous avons vu précédemment que la population de Vervins présente des effectifs efficaces particulièrement faibles comparés à ceux des autres populations étudiées et comparés aux effectifs classiquement observés dans les populations naturelles (ratio  $N_e/N = 0,01$  avec  $N_e = 40$  observé entre la population parentale et la G12). La valeur de  $N_e$  estimée entre la population initiale G0 et G2 est particulièrement faible ( $N_e = 9$ ). Cette observation, combinée avec celle de la réduction de diversité observée aux marqueurs microsatellites entre la population initiale G0 ( $H_e = 0,56$ ) et la génération 2 ( $H_e = 0,45$ ), suggère que cette population a subi un fort goulot d'étranglement dès son installation. Cette hypothèse avait déjà été soulevée au cours de la thèse de S. Paillard (2000) qui a suivi l'évolution de fréquences de gènes de résistance au mildiou, une maladie du blé, dans plusieurs populations de gestion dynamique dont celle de Vervins étudiée ici après 10 générations d'évolution. Dans son expérimentation, S. Paillard a montré que les gènes de résistance ont subi de forte variation de fréquences alléliques conduisant à leur élimination contrairement à ce qui était attendu pour cette population soumise à de fortes pressions parasitaires (Paillard *et al.* 2000). La reconstruction des haplotypes multilocus souligne bien le fait que la population semble s'être construite de façon particulière puisque 11 individus sur les 41 utilisés pour la reconstruction d'haplotype possèdent le même génotype multilocus à l'état homozygote. Cependant un grand nombre d'haplotypes sont retrouvés chez les autres individus à l'état hétérozygote en G2 de sorte que le nombre moyen d'allèles sur les 21 locus microsatellites ne diminue pas dans cette population contrairement à ce qui est attendu dans le cas d'un goulot d'étranglement (Charlesworth 2003). En revanche, comme attendu, les fréquences alléliques ont évolué très rapidement dès les premières générations, signe que cette population a subi d'importants phénomènes de dérive, conduisant à une population fortement différenciée des autres populations du réseau. Ce phénomène est accentué par l'apparition de nouveaux allèles spécifiques de cette population qui augmentent rapidement en fréquence.

La présence majoritaire d'un même génotype à l'ensemble des locus observés suggère que la population a subi un évènement démographique ou sélectif de forte ampleur à son installation de sorte que les individus de la population se sont trouvés en situation de très forte compétition à la première génération. Vingt ans après l'installation de cette population dans un lycée agricole, il est bien difficile d'identifier la cause précise de cette évolution particulière. Il est possible que les populations aient subies de très forts phénomènes de

sélection lors de leur installation dans les différents sites puisqu'elles étaient constituées de génotypes non spécifiquement adaptés aux différents milieux. En effet, les 16 génotypes parentaux utilisés pour constituer la population initiale étaient en partie des lignées commerciales (6 sur les 16 lignées utilisées) susceptibles d'être cultivées sur l'ensemble du territoire français, mais également des géniteurs INRA issus de croisements interspécifiques et choisis pour leur niveau de résistance aux maladies mais ne présentant pas nécessairement un niveau de rendement performant. De plus, les croisements forcés effectués pour constituer la population initiale ont pu rompre les associations alléliques favorables des variétés commerciales conduisant à une population initiale de mauvais niveau agronomique pour certains caractères. Toutefois cette explication paraît insuffisante pour expliquer une aussi forte réduction d'effectif efficace dans la population de Vervins.

### III-5 Conclusions

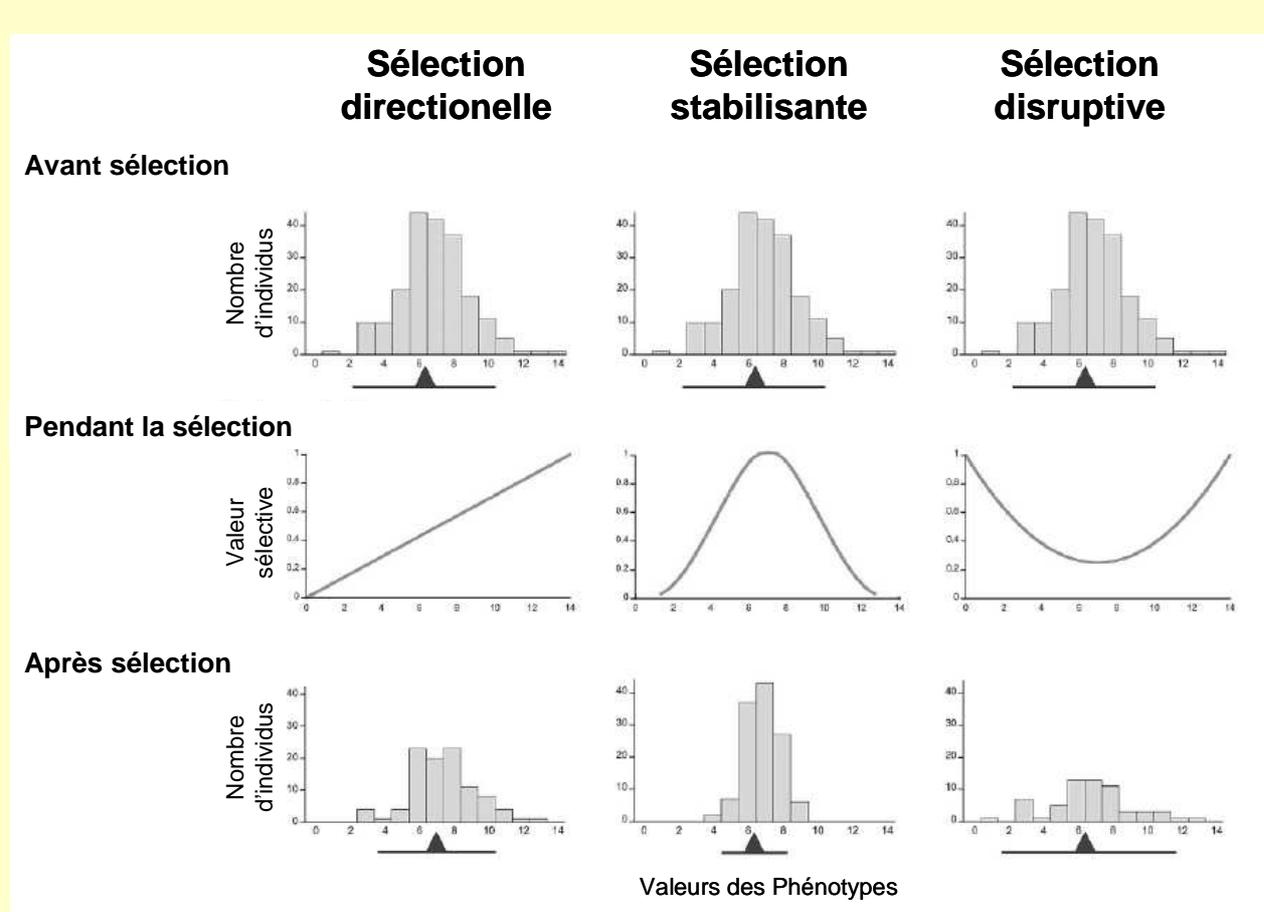
L'étude de l'évolution de marqueurs neutres dans les différentes populations et l'approche multilocus nous ont permis de clarifier certaines caractéristiques des populations. En premier lieu, le régime de reproduction mixte des populations semble jouer un rôle important dans l'évolution des populations. (i) La forte autogamie induit un fort déséquilibre de liaison à l'échelle d'un chromosome et une structuration en lignées à l'intérieur même des populations ; (ii) un taux d'allogamie résiduel permet l'apparition de nouvelles combinaisons alléliques et le maintien d'un certain niveau de diversité à l'intérieur des populations.

L'ensemble du génome est affecté par les forts effets combinés de la dérive et de la sélection mais ceux-ci semblent peu affecter la diversité intra population au bout de 12 générations d'évolution. A l'échelle du réseau de population de GD, la diversité inter populations est maintenue grâce à la différenciation observée entre populations du fait de la dérive et de l'isolement des populations. Le maintien de la diversité microsatellite est également liée au fort taux de mutation affectant ce type de marqueurs mais est aussi à mettre en relation avec des événements de contamination de la population initiale qui ont un impact sur la diversité de l'ensemble du génome. Enfin, l'étude de l'évolution de la diversité microsatellite nous a permis de détecter un épisode de goulot d'étranglement lors de l'installation de la population de Vervins pour laquelle des phénomènes de dérive importants sont observés. L'ensemble de ces observations est susceptible d'influer sur la réponse à la sélection dans les populations et sera donc repris dans les chapitres suivant visant à détecter la sélection au niveau phénotypique pour le caractère de précocité (Chapitre IV) et au niveau des gènes impliqués dans ce caractère (Chapitre V).

## *Chapitre 4*

### *EVOLUTION DES CARACTERES QUANTITATIFS : PRECOCITE DE FLORAISON ET COMPOSANTES DE LA VALEUR SELECTIVE*

Encadré IV-1 : Principales formes de sélection affectant la distribution des caractères quantitatifs.



**Distribution du caractère avant sélection** (en haut), **valeur sélective en fonction des valeurs phénotypiques** ou « fonction de fitness » (au milieu) et **distribution du caractère après sélection**. Le triangle sous les histogrammes indique la moyenne de la population et la barre verticale la variation autour de la moyenne ( $\pm 2$  fois l'écart type).

*Tiré de Kingsolver et Pfennig (2007).*

## **Chapitre IV: Evolution des caractères quantitatifs: Précocité de floraison et composantes de la valeur sélective**

### **IV-1 Introduction**

Mesurer la sélection en conditions naturelles n'est pas aisé. D'une part, les causes de variation des distributions phénotypiques d'une population, outre la sélection, sont multiples : variations environnementales, évènements de migration, dérive génétique ; d'autre part, il convient de déterminer précisément sur quels caractères agit la sélection : morphologie, traits d'histoire de vie, *etc.* ; enfin, la sélection peut opérer de différentes manières (sélection directionnelle, stabilisante ou disruptive) chacune influant de façon différente sur la variabilité et les moyennes des caractères (Encadré IV-1).

Dans le dispositif de gestion dynamique, une même population initiale G<sub>0</sub> à base génétique large, issue de croisements de 16 lignées parentales a été distribuée dans un ensemble de sites aux conditions environnementales contrastées. L'évolution phénotypique des populations aux cours des générations a fait l'objet de différents travaux dont les résultats sont présentés dans l'Encadré IV-2. Il en ressort que les populations ont subi d'importants changements phénotypiques aux cours des générations tant pour des caractères morphologiques (augmentation généralisée de la hauteur moyenne des plantes) que pour les traits d'histoire de vie (gradient Nord/Sud de précocité, évolution différenciée des résistances aux maladies). La sélection naturelle en relation avec des facteurs environnementaux a souvent été invoquée pour expliquer ces évolutions. Par exemple, l'évolution du niveau de résistances aux maladies a été mise en relation avec l'évolution des pressions pathogènes des sites de culture (Paillard *et al.* 2000a ; Paillard *et al.* 2000b ; Rhoné *et al.* 2007). La floraison tardive des populations du Nord comparées à celles du Sud a été mise en relation avec les caractéristiques climatiques des sites de culture et notamment la température (Goldringer *et al.* 2006).

En dépit des effectifs efficaces faibles, estimés à partir des marqueurs moléculaires (*cf.* Chapitre III), le rôle possible de la dérive génétique dans la différenciation des populations pour le caractère de précocité n'a pas été considéré de façon approfondie. Dans le cas de populations structurées, l'influence de la sélection sur la structuration des populations peut

être inférée en comparant la différenciation entre des caractères potentiellement soumis à la sélection et la différenciation de marqueurs neutres. La différenciation estimée à partir de marqueurs neutres ( $F_{ST}$ , Wright 1952) est considérée comme l'hypothèse nulle pour la différenciation attendue d'un caractère quantitatif à déterminisme génétique additif du fait de la dérive génétique seule. La différenciation des populations, estimée à partir de la répartition de la variance génétique d'un caractère entre ces populations ( $Q_{ST}$ , Spitze 1993), reflète les effets combinés de la dérive et de la sélection naturelle agissant sur le caractère. La détection de la sélection par comparaison de  $Q_{ST}$  avec le  $F_{ST}$  dans des populations naturelles évoluant dans des environnements contrastés est désormais une approche couramment utilisée. Elle a été mise en oeuvre pour un grand nombre d'organismes (*cf.* méta-analyse de Leinonen *et al.* 2008) et l'ensemble de ces études met en évidence que la sélection est un processus très répandu dans la nature.

L'objectif de ce chapitre est d'identifier les forces évolutives influant sur la précocité de floraison dans des populations de gestion dynamique de blé qui évoluent dans des conditions climatiques contrastées. Pour les plantes annuelles telles que le blé, la précocité de floraison est un caractère adaptatif majeur car il permet à la plante d'ajuster son cycle de développement pour que la reproduction se produise au moment où les conditions environnementales sont optimales (Remington et Purugganan 2003 ; Roux *et al.* 2006). La précocité de floraison est un caractère complexe contrôlé à la fois par l'état intrinsèque de la plante et par les conditions environnementales dans lesquelles elle évolue, en particulier la température et la longueur du jour. Dans un premier temps, l'évaluation des caractères au champ dans les conditions du Moulon est considérée car ce traitement correspond à la date "normale" de semis (novembre) des populations de gestion dynamique dans les différents sites de cultures. En comparant la différenciation des populations observée pour la précocité à celle observée pour les marqueurs microsatellites supposés neutres, je vais montrer que la différenciation de ce caractère ne peut être expliquée par l'action de la dérive seule. Puis je vais chercher à quantifier la force de la sélection agissant directement sur le caractère en me basant sur la réponse à la sélection observée entre 2 générations (G2 et G12) sur le site du Moulon et en tenant compte des corrélations avec d'autres caractères potentiellement sélectionnés. L'évolution de la précocité de floraison observée dans les populations a souvent été mise en relation avec les conditions climatiques contrastées dans les différents sites, en particulier la durée des basses températures. Celle-ci étant plus longue au Nord qu'au Sud, elle peut conduire à un allongement du cycle végétatif. En utilisant l'approche de régression multiple décrite par Lande et Arnold (1983), les gradients et différentielles de sélection

attendus ont été estimés en génération 2 pour la précocité et les autres caractères dans des conditions de durée de vernalisation différentes (4 semaines et 8 semaines). Enfin, les stratégies de réponses aux facteurs environnementaux (sensibilité à la vernalisation et à la photopériode) sélectionnées dans les différentes populations seront étudiées en fonction des caractéristiques climatiques de chaque site.

**Encadré IV-2. Principaux résultats des travaux antérieurs portant sur l'évolution des caractères quantitatifs dans les populations PA du réseau de gestion dynamique.**

<p><b>Thèse J. David (1992)</b></p>	<p><b>Populations et générations étudiées :</b> Populations de Châlons-sur-Marne, Le Moulon, Montreuil, Nérac, Rennes, Toulouse, Vervins en génération 6.</p> <p><b>Caractères :</b> Précocité et hauteur évaluées sur le site du Moulon et de Rennes. Productivité à l'échelle d'une grande parcelle évaluée sur les sites du Moulon et Clermont-Ferrand.</p> <p><b>Principaux résultats :</b> - Les populations fleurissent selon un gradient Nord-Sud de précocité en relation avec les caractéristiques climatiques des différents sites de culture. - La hauteur moyenne des plantes augmente de façon unidirectionnelle dans toutes les populations du fait de la compétition entre plantes au sein de chaque population. - Il n'y a pas d'interaction entre populations et sites de culture pour la productivité estimée en grandes parcelles malgré des conditions environnementales très distinctes entre les sites du Moulon et de Clermont-Ferrand, ce qui suggère une stabilité de la productivité lorsque les populations sont en conditions de compétition.</p> <p>➔ Il existe une sélection pour l'adaptation locale et pour l'aptitude à la compétition. ➔ Il existe une relation antagoniste entre valeur sélective individuelle et valeur sélective de la population, de sorte que les plantes hautes produisant moins de graines mais présentant une bonne aptitude à la compétition sont favorisées dans l'ensemble des populations.</p>
<p><b>Thèse S. Paillard (1999)</b></p>	<p><b>Populations et générations étudiées :</b> Populations de Châlons-sur-Marne, Le Moulon, Montreuil, Rennes, Toulouse, Vervins en génération 10.</p> <p><b>Caractères :</b> Résistances aux maladies (oïdium, rouilles jaune et brune et piétin verse) évaluées sur les sites du Moulon, Montpellier, Rennes.</p> <p><b>Principaux résultats :</b> ➔ Il existe une adaptation locale aux pressions parasitaires : le niveau de résistance évolue dans les différentes populations en relation avec les pressions parasitaires de chaque site</p>

## IV-2 Méthodes

L'ensemble des évaluations effectuées sur les populations étudiées est détaillé dans le chapitre 2, « matériels et méthodes », le Tableau IV-1 présente une synthèse des mesures disponibles pour analyser l'évolution phénotypique des populations.

Tableau IV-1.

	Caractère évolué et conditions	Dates et matériel évalué	Mesures
<b>HdCHP</b>	<b>Précocité au champ</b> en conditions "normales" de semis des populations de GD (Semis au champ en automne)	Printemps 2005 : Ensemble des populations étudiées (semis Nov. 2004) Printemps 2006 : Ensemble des populations étudiées (semis Nov. 2005)	Notation de la date d'épison moyenne par ligne de 20 plantes d'un même génotype, 2 répétitions spatiales Notation de la date d'épison moyenne par ligne de 20 plantes d'un même génotype.
<b>HT-CHP</b> <b>NE-CHP</b> <b>NGP-CHP</b> <b>PGP-CHP</b>	<b>Hauteur, nombre d'épis, nombre de grains et poids de grains</b> en conditions "normales" de semis des populations de GD (Semis au champ en automne)	Été 2005 : Ensemble des populations étudiées	Notation de 3 plantes par génotype prises au hasard dans la ligne de 20 plantes évaluées pour la précocité sur une des 2 répétitions spatiales
<b>Hd8S</b> <b>Hd4S</b> <b>HdNV</b>	<b>Précocité</b> en conditions variables de vernalisation en pépinière. Transplantation de plantules totalement vernalisées (8S), partiellement vernalisées (4S) et non vernalisées (NV) en Avril	2003 : 16 génotypes parentaux 2005 : Ensemble des populations étudiées 2007 : 16 génotypes parentaux	Notation de 5 à 7 plantes par génotype (selon les traitements) semées sur une ligne Notation de 3 à 6 plantes par génotype (selon les traitements) semés sur une ligne Notation de 3 plantes par génotypes semées sur une ligne, 2 répétitions spatiales
<b>HT-4S</b> <b>NE-4S</b> <b>NG-4S</b> <b>PG-4S</b> <b>HT-8S</b> <b>NE-8S</b> <b>NGP-8S</b> <b>PGP-8S</b>	<b>Hauteur, nombre d'épis, nombre de grains et poids de grains</b> de plantes ayant subi des conditions variables de vernalisation (4S et 8S) au stade plantule et transplantées en Avril en pépinière	Été 2005 : Ensemble des populations étudiées Été 2007 : 16 génotypes parentaux	Notation des 3 plantes par génotype semées sur une ligne et évaluées précédemment pour la précocité Notation des 3 plantes par génotype semées sur une ligne et évaluées précédemment pour la précocité, 2 répétitions spatiales
<b>HdJC</b>	<b>Précocité</b> de plantes totalement vernalisées en condition de jours courts en chambre de culture	2003: 16 génotypes parentaux 2006: Ensemble de la population	Notation de 3 plantes par génotype repiquées dans un pot, 2 répétitions spatiales Notation de 2 plantes par génotypes repiquées dans un pot

Rappel des abréviations: Hd=date d'épison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante.

## IV-2.1 Estimation des variances environnementales

Pour un caractère quantitatif, la valeur phénotypique (P) mesurée pour un individu est fonction de 2 facteurs : la valeur du génotype (G) de l'individu et l'ensemble des causes non génétiques et non contrôlées de variations dites « environnementales » (E). La valeur phénotypique d'un individu est exprimée par :  $P = \mu + G + E$ , avec  $\mu$  moyenne de la population. Dans une population constituée d'individus de plusieurs génotypes, en l'absence d'interaction entre les génotypes et l'environnement ou dans un environnement donné, la variance phénotypique peut alors s'écrire:  $V_P = V_G + V_E$ . Pour caractériser l'évolution des populations il est nécessaire d'avoir accès à la variance génétique. L'estimation de cette variance est souvent obtenue à partir de la décomposition de la variance phénotypique ( $V_G = V_P - V_E$ ) et nécessite donc une bonne estimation de la variance environnementale.

Pour une population de génotypes répétables (*i.e.* génotypes homozygotes à tous les locus, comme des lignées ou des clones), la variance environnementale peut être estimée à partir de l'observation de plusieurs individus d'un même génotype. Dans les expérimentations menées sur les populations de gestion dynamique pour cette étude, l'évaluation phénotypique des différents caractères a été effectuée sur plusieurs descendants d'autofécondation de génotypes issus des différentes populations pour lesquelles les taux d'autofécondation sont très élevés (*cf.* Chapitre III). Ainsi la variance observée intra génotype sera considérée comme une variance d'origine environnementale ( $V_E$ ) tandis que la variance inter génotypes correspondra à la variance génétique totale ( $V_G$ ) estimée dans chaque population.

L'ensemble des mesures de précocité au champ a été effectué 2 années consécutives avec 2 répétitions spatiales la première année d'évaluation (Tableau IV-1). Ce dispositif nous permet d'avoir une estimation de la variance environnementale inter annuelle et spatiale pour ce caractère et donc une bonne estimation de  $V_G$ . Les autres caractères évalués au champ ou en pépinière ont été mesurés pour l'ensemble des populations sur 3 ou 5 plantes (selon le dispositif) par génotype sur une ligne, une seule année (Tableau IV-1). Notons qu'un dispositif dans lequel les 3 à 5 plantes répétées par génotype auraient été installées dans différents blocs randomisés sur l'ensemble de la parcelle aurait permis une meilleure estimation de la variance environnementale. Cependant, du fait du grand nombre de plantes manipulé pour chaque évaluation phénotypique, un tel dispositif aurait été techniquement difficile à mettre en œuvre et aurait pu conduire à une augmentation de l'erreur lors de la mise

en place de l'expérimentation (multiplication des manipulations de sachets de graines, nécessité d'identifier les génotypes plante à plante) ou lors de l'acquisition des données.

La variance environnementale estimée à partir des répétitions de 3 à 5 plantes sur une ligne reflète une variance intra micro-parcellaire ( $V_{E\mu\text{Parcelle}}$ ) qui peut être sous-estimée comparée à la variance environnementale à l'échelle de l'expérimentation. Pour palier à ce problème d'estimation de la variance environnementale, les 16 génotypes parentaux ont été replantés plusieurs années pour tenter d'obtenir une bonne estimation de la variance environnementale que l'on peut ensuite extraire de la variance phénotypique estimée dans les populations en 2005. Ceci nécessite cependant que les variances environnementales estimées les différentes années soient comparables à la variance environnementale estimée en 2005, année d'évaluation de l'ensemble des populations.

Dans un premier temps,  $V_E$  a été estimée à partir des 16 lignées parentales qui ont été répétées à plusieurs reprises (répétitions spatiales ou annuelles selon les caractères, Tableau IV-1) pour chaque année d'évaluation (modèle (1)) et en tenant compte de l'ensemble des années (modèle (2)).

$$Y_{ij} = \mu + geno_i + \varepsilon_{ij} \quad (1)$$

avec  $Y_{ij}$  mesure sur la plante  $j$  du génotype  $i$ ,

$geno_i$  effet du génotype  $i$ ,

$\varepsilon_{ij}$  terme résiduel de variance  $V_{E\mu\text{Parcelle}}$ .

Les  $V_{E\mu\text{Parcelle}}$  ont ainsi été estimées pour chaque année d'évaluation (par exemple, pour le caractère Hd8S, on dispose d'une  $V_{E\mu\text{Parcelle}}$  pour l'année 2003, pour l'année 2005 et pour l'année 2007 ; pour l'année 2007 un effet « rep » a été déclaré pour tenir compte de la répétition spatiale (modèle (1'))).

Puis ces variances ont été comparées entre-elles et à la variance environnementale ( $V_{E\text{Parcelle}}$ ) estimée à partir des 16 génotypes parentaux évalués plusieurs années selon les caractères (Tableau IV-1), en se basant sur le modèle d'analyse de la variance suivant :

$$Y_{ijk} = \mu + geno_i + an_j + \varepsilon_{ijk} \quad (2)$$

avec  $Y_{ijk}$  la mesure sur la plante  $k$  de la répétition  $j$  du génotype  $i$ ,

$geno_i$  l'effet du génotype  $i$ ,

$an_j$  l'effet de l'année (Par exemple, pour Hd8S,  $i = 2003, 2005$  et  $2007$ . Pour l'année 2007, comme l'effet  $rep$  n'est pas été déclaré, car non significatif dans le modèle (1'))

$\varepsilon_{ij}$  le terme résiduel de variance  $V_{E\text{Parcelle}}$ .

Les composantes de la variance ont été estimées à l'aide de la procédure VARCOMP du logiciel SAS version 8 (méthode REML, Restricted Maximum Likelihood).

Etant données les conditions météorologiques particulières de l'été 2007 (très forte pluviométrie, faible ensoleillement, températures estivales très en dessous des valeurs normales), l'expérimentation effectuée en pépinière sur les génotypes parentaux en 2007 pour estimer la précocité de floraison et les composantes de la valeur sélective en conditions de vernalisation partielle (4S) ou complète (8S) ont conduit à l'estimation d'une variance environnementale très élevée pour la plupart des caractères. Elle ne peut donc être directement comparée à la variance environnementale estimée les autres années (Tableau IV-2).

Tableau IV-2 : Estimation des variances environnementales pour les 16 lignées parentales évaluées à partir de 3 à 5 plantes disposées sur une ligne plusieurs années.

		<b>Hd8S</b>		<b>Hd4S</b>	
		$V_E$	s.d.	$V_E$	s.d.
Modele (1)	2003	572.3	93.3	1992.9	313.1
	2005	808.3	208.4	5573.3	1513.9
	2007	3242.1	533.4	21860.8	4381.5
Model (2)	2003-2005-2007	1741.0	182.8	8949.9	1002.7
	2003-2005	643.3	89.2	2894.2	394.3

		<b>HT-8S</b>		<b>NE-8S</b>		<b>NG-8S</b>		<b>PG-8S</b>	
		$V_E$	s.d.	$V_E$	s.d.	$V_E$	s.d.	$V_E$	s.d.
Modele (1)	2005	21.9	5.7	6.6	1.7	13094.7	3315.4	13.1	3.3
	2007	43.1	6.9	5.1	0.8	10507.2	1770.1	139.4	23.6
Model (2)	2005-2007	38.0	5.2	6.7	0.9	11454.5	1606.3	117.4	16.1

		<b>H4S</b>		<b>NE-4S</b>		<b>NG-4S</b>		<b>PG-4S</b>	
		$V_E$	s.d.	$V_E$	s.d.	$V_E$	s.d.	$V_E$	s.d.
Modele (1)	2005	95.9	25.0	14.5	3.8	20479.6	5352.1	19.4	5.1
	2007	61.3	12.1	12.8	2.5	18702.9	3721.4	116.4	23.6
Model (2)	2005-2007	75.0	10.9	13.7	2.0	19552.3	3092.1	78.2	12.3

La comparaison des différentes variances environnementales (Tableau IV-2) montre qu'elles varient de façon importante d'une année sur l'autre. Cette observation rend délicate l'utilisation de la variance environnementale estimée en combinant l'ensemble des années d'évaluation sur les 16 génotypes parentaux (modèle (2)) pour extraire la variance génétique

des populations, toutes évaluées en 2005. De plus, la variance estimée en 2005 à partir des génotypes parentaux, ne semble pas sous-estimée par rapport aux autres années. Par ailleurs les effets de répétitions spatiales, lorsqu'ils sont estimables (évaluation 2007), ne sont pas significatifs, ce qui suggère que la variance à l'échelle parcellaire n'est pas plus importante que la variance estimée à l'échelle micro-parcellaire. Par la suite, il semble donc plus judicieux d'extraire les variances génétiques pour ces caractères par la décomposition phénotypique, à partir de la variance environnementale micro-parcellaire, estimée à l'intérieur de chaque population évaluée en 2005 et en se basant sur les répétitions des 3 à 5 plantes sur une ligne.

## **IV-2.2 Evolution des caractères au cours des générations dans chaque population**

### **IV-2.2.1 Evolution phénotypique des caractères**

Les valeurs moyennes des caractères dans les différentes populations ont été estimées à partir des moyennes calculées par génotype sur les 3 à 5 plantes évaluées. Pour le caractère de précocité de floraison estimé au champ 2 années consécutives, les valeurs moyennes par génotype ont été obtenues à l'aide de la fonction *LS-Means* à la suite d'une analyse de variance dans laquelle un effet génotype et un effet année ont été déclarés. Cette fonction permet d'obtenir des moyennes ajustées pour les effets significatifs déclarés dans le modèle (SAS OnlineDoc V8).

Pour tous les caractères, les moyennes phénotypiques des caractères ont été comparées entre populations et entre générations dans une même analyse de variance en utilisant la procédure GLM (Generalized Linear Model, procédure conseillée pour traiter les analyses de la variance lorsque les classes sont déséquilibrées, SAS OnlineDoc v8) du logiciel SAS suivant le modèle suivant :

$$Y_{ijk} = \mu + pop_i + gene(pop)_{ij} + \varepsilon_{ijk} \quad (3)$$

avec  $Y_{ijk}$  la moyenne du génotype  $k$  sur 2 années (pour HdCHP) ou sur les 3 plantes (pour les autres caractères),

$pop_i$  l'effet de la population ( $i =$  Moulon, Toulouse ou Vervins),

$gene(pop)_{ij}$  l'effet de la génération ( $j = G2, G7$  ou  $G12$ ) dans la population  $i$  considérée,

$\varepsilon_{ijk}$  terme résiduel intra population, intra génération.

Dans cette analyse, tous les effets sont considérés comme fixes car chaque population peut être considérée comme résultant d'une évolution unique obtenue à partir de la population initiale G0 à la suite d'évènements historiques particuliers. Les différences de moyennes des différentes populations considérées dans l'analyse ont été testées 2 à 2 avec un test de comparaison multiple de Student-Newman-Keuls (SNK).

#### **IV-2.2.2 Caractérisation de la variabilité génétique intra population des caractères**

Dans chaque population, à chaque génération, les composantes de la variance ont été estimées à partir des mesures plante à plante selon 2 modèles :

Pour la précocité de floraison évaluée au champ (HdCHP) :

$$Y_{ij} = \mu + geno_i + an_j + \varepsilon_{ij} \quad (4)$$

avec  $Y_{ijk}$  moyenne sur une ligne du génotype  $i$  l'année  $j$

$geno_i$  effet aléatoire du génotype  $i$  ( $i = 1, \dots, 60$ ) de variance  $V_G$

$an_j$  effet fixe de l'année d'évaluation ( $j = 2005$  ou  $2006$ ). Pour simplifier l'analyse, la valeur moyenne de précocité entre les 2 répétitions spatiales de l'année 2005 est considérée (l'effet *répétition* en 2005 étant non significatif dans l'ensemble des analyses effectuées en préliminaire).

$\varepsilon_{ij}$  terme résiduel de variance  $V_E$

Pour les autres caractères évalués sur 3 plantes sur une ligne au champ (HT-CHP, NE-CHP, etc.) ou en pépinière, la décomposition de la variance a été effectuée dans chacune des populations en se basant sur le modèle d'analyse (1) présenté précédemment pour extraire  $V_G$  et  $V_E$

La décomposition de la variance en facteurs génétiques et environnementaux nous permet d'estimer des paramètres rendant compte de la variabilité génétique des populations pour le caractère considéré :

- Héritabilité : part de la variance phénotypique d'origine génétique  $h^2 = V_G/V_P$  .

Cette mesure de diversité est couramment employée pour mesurer la diversité génétique d'un caractère quantitatif sur laquelle va pouvoir agir la sélection dans une population. Cependant

l'emploi de cette mesure pour estimer la capacité d'une population à répondre à la sélection est critiqué car elle n'est valable que dans un environnement donné et pour une population donnée et n'est donc pas extrapolable à d'autres populations ou environnements. Les héritabilités sont considérées comme significativement différentes de zéro lorsque l'effet « *geno* » est significatif dans le modèle d'analyse.

- Coefficient de variation génétique :  $CV_G = \sqrt{V_G} / \mu$  (avec  $\mu$  la moyenne de la population).

Constatant que certains caractères présentent une héritabilité faible, résultant d'une variance environnementale élevée, en dépit d'une variance génétique forte, Houle (1992) propose d'utiliser cette mesure comme alternative à l'héritabilité pour rendre compte de la diversité d'un caractère à l'intérieur d'une population, relativement à la moyenne du caractère. Le coefficient de variation environnementale ( $CV_E$ ) a également été estimé.

#### **IV-2.3 Caractérisation de la variabilité génétique inter populations des caractères : mise en évidence de la sélection.**

De la même façon que pour les marqueurs génétiques, la structuration d'un caractère quantitatif entre populations peut être estimée à partir d'une mesure de différenciation ( $Q_{ST}$ ) qui représente la proportion de la variance génétique totale observée entre les populations. La structuration de la variabilité génétique entre populations est estimée par l'expression suivante :

$$Q_{ST} = \frac{(1 + F_{IS})V_{pop}}{(1 + F_{IS})V_{pop} + 2V_{geno}} \quad (\text{Spitze 1993})$$

avec  $V_{pop}$  la variance inter-population,  $V_{geno}$  la variance génétique intra-population et  $F_{IS}$  le coefficient de consanguinité. Dans le cas d'espèces autogames et de l'étude de leurs descendance issues d'autofécondations contrôlées ( $F_{IS} \approx 1$ ), cette expression devient :

$$Q_{ST} = \frac{V_{pop}}{V_{pop} + V_{geno}}$$

Pour la précocité de floraison évaluée au champ (HdCHP), la variance inter populations et la variance intra population ont été estimées à partir du modèle d'analyse suivant :

$$Y_{ijk} = \mu + pop_i + geno(pop)_{ij} + an_k + \varepsilon_{ijk} \quad (5)$$

avec  $pop_i$  l'effet de la population  $i$  (Moulon, Toulouse ou Vervins) de variance  $V_{pop}$ ,

$geno(pop)_{ij}$  l'effet du génotype  $j$  ( $j = 1 \dots 60$ ) à l'intérieur de la population  $i$  de variance  $V_{geno}$

$an_k$  l'effet fixe de l'année d'évaluation ( $k = 2005$  ou  $2006$ ),

$\varepsilon_{ijk}$  le terme résiduel.

Les effets  $pop$  et  $geno$  sont considérés ici comme aléatoires puisque les populations étudiées peuvent être considérées ici comme un échantillon aléatoire représentatif des populations issues de la population initiale  $G_0$ . Pour les autres caractères évalués au champ ou en pépinière, le modèle d'analyse considéré pour extraire les variances inter et intra population est le même sans déclarer déclaration d'effet « année » (modèle (5')). Les valeurs de  $Q_{ST}$  obtenues sont considérées comme significatives lorsque l'effet "population" est significatif dans le modèle d'analyse de la variance.

La différenciation estimée à partir des caractères quantitatifs a ensuite été confrontée à la différenciation observée à partir des 21 locus microsatellites ( $F_{ST}$  cf. Chapitre III) supposés neutres, pour mettre en évidence le rôle de la sélection dans la divergence des populations.

#### **IV-2.4 Estimation de la sélection directionnelle directe et indirecte : approche multivariée**

##### **IV-2.4.1 Estimation des gradients de sélection directionnelle observés dans les conditions "champ" au Moulon.**

L'approche utilisée ici est celle développée dans le cadre de la thèse de J. David (1992). Les gradients de sélection sont estimés en se basant sur la réponse à la sélection évaluée sur le site du Moulon entre 2 générations étudiées et en tenant compte des corrélations entre caractères contribuant plus ou moins à la valeur sélective des individus. Les gradients de sélection ainsi estimés, dans le contexte environnemental du Moulon, correspondent aux forces de sélection agissant directement sur le caractère à chaque génération et ayant conduit à la différenciation observée des populations sur le site du Moulon, au bout des  $N$  générations considérées. Il s'agit ici de déterminer si la précocité de floraison est sélectionnée de façon directe ou indirecte. Du fait de l'évolution continue du caractère de précocité dans les populations

considérées jusqu'en génération 12, notamment dans les populations les plus extrêmes (Vervins et Toulouse), seul les gradients de sélection directionnelle ont été considérés.

Dans le dispositif de gestion dynamique, les gradients de sélection directionnelle peuvent être estimés à partir de l'expression de la réponse à la sélection  $\Delta z$  :

$$\Delta z = \mathbf{G}\mathbf{P}^{-1}s = \mathbf{G}\beta \quad (\text{Lande 1979})$$

avec  $\mathbf{G}$  la matrice des variances covariances génétiques,  $\mathbf{P}$  la matrice des variances covariances phénotypiques,  $s$  la différentielle de sélection (Par définition,  $s = \bar{z}^* - \bar{z}$  avec  $\bar{z}$  vecteur des moyennes des  $n$  caractères dans la population et  $\bar{z}^*$  vecteur des moyennes après sélection, c'est-à-dire moyennes des parents sélectionnés) et  $\beta$  le vecteur des gradients de sélection directionnelle. Cette expression est la version multivariée de l'équation dite « du sélectionneur », donnant la relation entre réponse à la sélection  $R$  et héritabilité  $h^2$  :  $R = h^2s$ .

Le gradient de sélection directionnelle annuel  $\beta_i$  appliqué à la population de départ pour aboutir en  $N$  générations à la population  $i$  dans les conditions du Moulon peut alors s'écrire :

$$\beta_i = \mathbf{G}_0^{-1} (z_i - z_0) / N$$

avec  $z_0$  ( $z_i$ ) vecteur des moyennes dans la population initiale considérée (dans la population  $i$  à la génération observée). La population initiale présentant un effectif d'échantillonnage plus faible ( $n = 35$ ) que les autres populations ( $n \approx 55$ ), cette approche a été effectuée en se basant sur la réponse à la sélection observée entre la génération 2 et la génération 12 pour chacune des populations. De plus, cette approche suppose que la matrice de variance-covariance évolue peu du fait de la sélection ou de la dérive au cours des générations. Ceci semble être le cas dans les populations considérées puisqu'il existe un fort coefficient de corrélation entre les termes de la matrice de variance-covariance estimés en G2 et ceux estimés en G12 (LM :  $R = 0,86$  ; TO :  $R = 0,85$  et VV :  $R = 0,79$ ). Enfin cette approche suppose la multinormalité des variables avant sélection.

Les caractères considérés dans cette approche sont la date d'épiaison (HdCHP), la hauteur à maturité (HT-CHP), le nombre d'épis (NE-CHP), le nombre de grains par épi (NE\_G-CHP) et le poids de 1000 grains (P1000G-CHP). Le nombre de grains par plante et le poids de grains par plante étant très corrélés entre eux et également très corrélés au nombre d'épis par plante ( $R > 0,75$ ), ces 2 caractères n'ont pas été considérés pour ne pas déséquilibrer la matrice de

variance covariance (Fox *et al.* 2001). Les moyennes de ces différents caractères estimées à la génération 2 et à la génération finale observée G12 ont été réduites par l'écart-type de la population initiale. De cette façon, les différences entre la génération 2 et la génération 12 dans les différentes populations sont exprimées en unité d'écart-type de la génération 2.

La matrice de variance-covariance génétique  $\mathbf{G}_0$  pour la population initiale considérée (génération 2 de chaque population) a été calculée en se basant sur un modèle d'analyse de variance multivarié en utilisant l'option MANOVA de la procédure GLM (logiciel SAS).  $\mathbf{G}_0$  a été obtenue par soustraction de la matrice des carrés et produits moyens résiduels à la matrice des carrés et produits moyens de l'effet génotype.

#### **IV-2.4.2 Estimation des gradients de sélection attendus dans des contextes variant pour les besoins en vernalisation**

La sélection phénotypique sur des populations dans un environnement donné peut être analysée par les techniques de régression multiples décrites par Lande et Arnold (1983). Dans cette approche, la différentielle de sélection  $s$  est vue comme la covariance entre le phénotype et la valeur sélective relative :  $s = Cov[w, z]$ . Le gradient de sélection, défini par  $\beta = \mathbf{P}^{-1}s$ , est le vecteur des coefficients de régression de la valeur sélective relative sur les caractères phénotypiques. La différentielle de sélection quantifie la force de sélection directionnelle globale incluant les effets indirects issus des corrélations entre caractères, tandis que le gradient de sélection mesure la force de sélection directionnelle agissant directement sur le caractère. La comparaison de ces 2 paramètres est instructive car un caractère peut montrer une réponse à la sélection sans être soumis directement à la sélection. Dans ce cas, la différentielle de sélection sera significativement différente de 0, alors que le coefficient de régression partielle correspondant sera non significatif (Lynch et Walsh 1998).

Dans cette analyse, le poids de grains total par plante (PGP) a été retenu comme mesure de la valeur sélective car cette variable représente l'investissement des plantes dans la reproduction sexuée (PGP est très corrélé au nombre de grains total par plante NGP composante essentielle du succès reproducteur) tout en intégrant une part de viabilité des descendants futurs. Dans chaque population considérée, la valeur sélective relative d'un génotype ( $w_{PGP}$ ) a été obtenue en divisant PGP par la valeur moyenne de la population ( $w_{PGP} = PGP / moy(PGP)$ ). Toutes les variables explicatives utilisées dans les analyses de régression (date d'épiaison Hd, hauteur à maturité HT, nombre d'épis NE, nombre de grains par épi NE\_G, poids de 1000 grains

P1000G) ont été centrées et réduites (moyennes nulles et variances unitaires). Ceci fournit des différentielles et des gradients de sélection normés ce qui permet une comparaison directe des coefficients de sélection entre caractères et entre populations (Lande et Arnold 1983 ; Kingsolver et Pfennig 2007).

La covariance entre valeur sélective relative et les différents caractères a été obtenue à l'aide de la procédure CORR du logiciel SAS. La différentielle est considérée comme significative si le coefficient de corrélation entre la valeur sélective et le caractère est significatif. Les gradients de sélection ont été estimés avec la procédure REG dans SAS.

## IV-3. Résultats

L'ensemble des moyennes des caractères par populations est donnée en Annexe 3.

### IV-3.1 Evolution phénotypique de la précocité et des composantes de la valeur sélective au champ

La précocité de floraison des populations de gestion dynamique évaluée au champ sur un même site montre que les populations évoluent de façon divergente pour ce caractère (Figure IV-1). Dès la première génération observée (G2), les populations du Moulon et de Vervins présentent une floraison plus tardive en moyenne que les populations parentale et initiale (G0) ainsi que la population de Toulouse. Au cours des générations, la distribution des valeurs de précocité dans la population de Vervins tend à se décaler vers des valeurs de plus en plus tardives tandis que dans la population de Toulouse, la distribution du caractère se décale vers des valeurs plus précoces. L'analyse de la variance (Tableau IV-3) montre pour ce caractère que les effets spatiaux (entre populations dans des sites différents) sont plus importants que les effets temporels (entre générations à l'intérieur d'une population).

Comparée aux autres populations et à la distribution générale du caractère, la population de Vervins présente une distribution asymétrique qui s'écarte de la normalité en G2 avec une part de la population épiant autour de 1200°J tandis qu'une autre partie épie autour de 1300°J. L'étude des marqueurs microsatellites et la reconstruction d'haplotypes multilocus en particulier (*cf.* Chapitre III) ont mis en évidence que cette population a subi un fort goulot d'étranglement lors de son implantation et est composée à la fois de génotypes très divers sur l'ensemble des locus ( $\approx 2/3$ ) et de génotypes présentant la même composition allélique à tous les locus ( $\approx 1/3$ ). Douze génotypes, probablement issus d'une même lignée car identiques aux 14 locus microsatellites servant à la reconstruction d'haplotypes, ont une moyenne de précocité de 1277°J (valeurs comprises entre 1149°J et 1306°J). Par ailleurs, 8 autres individus ayant des haplotypes multilocus très proches (un seul allèle différent sur les 14 locus par rapport au génotype évoqué précédemment) présentent une moyenne de 1278°J (valeurs comprises entre 1212°J et 1306°J). Ces génotypes très proches viennent gonfler la classe de génotypes tardifs. Ainsi, la distribution en 2 classes observée dans cette population en G2 s'explique par la présence de génotypes identiques ou fortement apparentés épiant tardivement autour de 1300°J (1/3 des génotypes de la population) tandis que l'autre partie de la population épie plus précocement, autour de 1200°J.

Parmi les autres caractères évalués dans cette expérimentation, la hauteur est le caractère qui évolue le plus avec une augmentation généralisée de la taille dans toutes les populations. Cette évolution est particulièrement importante entre la génération 2 et la génération 7 notamment dans la population de Vervins. Malgré cette évolution unidirectionnelle affectant l'ensemble des populations, l'effet inter populations reste important du fait de la taille significativement plus haute de la population de Vervins comparée aux autres populations. Les composantes de la valeur sélective évoluent de façon plus erratique au cours des générations dans chacune des populations mais tendent à augmenter comparées à la population initiale. Les caractères présentant le plus de différences entre les populations en génération 12 sont le nombre de grains par épi, plus élevé dans les populations du Nord qu'à Toulouse, et le poids de 1000 grains, dont la valeur moyenne est particulièrement élevée dans la population de Vervins. Cependant la productivité globale par plante (nombre et poids de grains par plante) est du même ordre de grandeur dans les 3 populations, la population de Toulouse compensant sa faible productivité par épi par un plus grand nombre d'épis par plante.

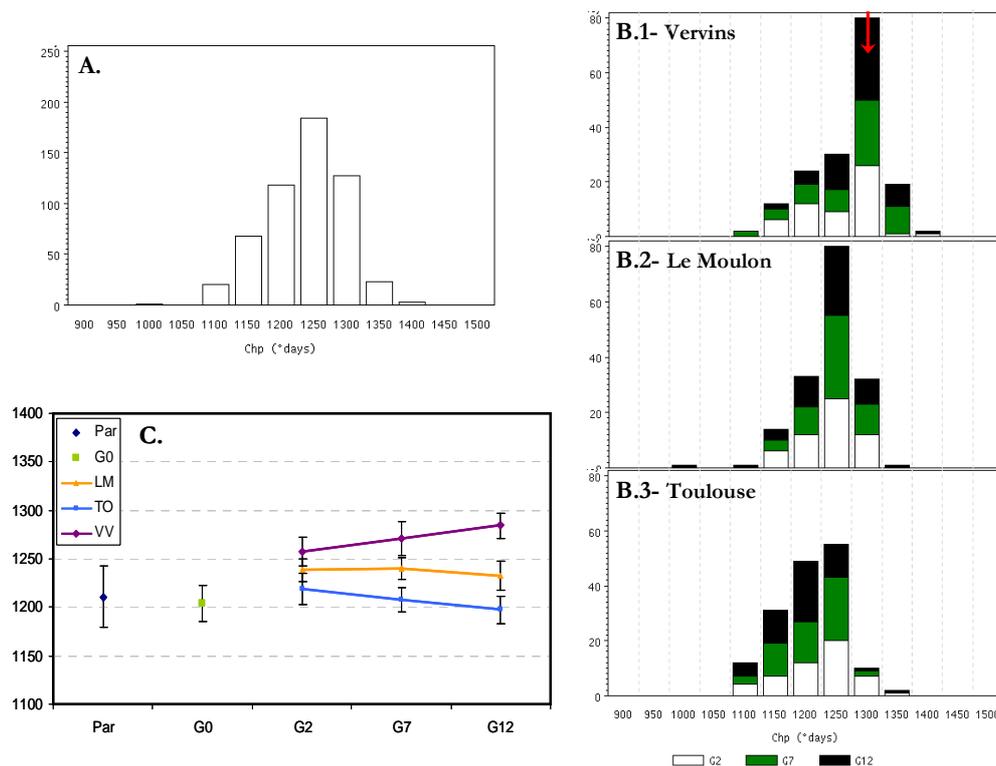


Figure IV-1 : Evolution de la précocité de floraison évaluée au champ dans les différentes populations aux cours des générations. A. Distribution du caractère toutes populations confondues ; B. Distribution du caractère dans les 3 populations étudiées et évolution au cours des générations ; C. Evolution des moyennes par population au cours des générations. Les bars d'erreurs correspondent aux intervalles de confiance à 95% estimés sur la moyenne. La flèche rouge indique la valeur moyenne des génotypes issus d'une même lignée identifiés par l'étude des marqueurs microsatellites dans la population de Vervins en génération 2.

Tableau IV-3 : Moyennes, écarts-types (en italique) et valeurs des *F*-ratios et *P*-values associées de l'analyse de la variance suivant le modèle (3) décrit dans la section « IV-2. Méthodes » pour les caractères évalués au champ dans la populations parentale, la population initiale et les 3 populations étudiées en génération 2 et en génération 12 (VV : Vervins ; LM : Le Moulon ; TO : Toulouse).

	P.Parentale (n=16)	P.Initiale G0 (n=35)	VV-G2 (n=55)	VV-G12 (n=59)	LM-G2 (n=)	LM-G12 (n=54)	TO-G2 (n=)	TO-G12 (n=53)	<i>Pop</i>		<i>Gene(Pop)</i>	
									<i>F</i>	<i>P</i>	<i>F</i>	<i>P</i>
HdCHP (°J)	1210.9 <i>60.1</i>	1203.7 <i>54.3</i>	1257.4 <i>55.5</i>	1284.4 <i>50.7</i>	1238.5 <i>44.4</i>	1232.5 <i>54.1</i>	1219.0 <i>58.2</i>	1197.4 <i>50.1</i>	60.4	<0.0001	2.2	0.0461
HT-CHP (cm)	87.3 <i>17.2</i>	89.2 <i>11.9</i>	98.7 <i>14.1</i>	118.6 <i>11.6</i>	91.7 <i>15.0</i>	104.3 <i>9.4</i>	93.3 <i>11.5</i>	101.3 <i>13.4</i>	45.1	<0.0001	24.2	<0.0001
NE-CHP	4.5 <i>1.4</i>	5.1 <i>1.3</i>	4.9 <i>1.4</i>	4.8 <i>1.1</i>	5.2 <i>1.4</i>	5.2 <i>1.5</i>	5.2 <i>1.4</i>	5.4 <i>1.4</i>	3.5	0.0315	1.3	0.2689
NGP-CHP	187.9 <i>53.4</i>	221.3 <i>62.9</i>	250.2 <i>72.5</i>	240.7 <i>67.9</i>	235.5 <i>70.8</i>	255.6 <i>85.1</i>	243.9 <i>70.4</i>	235.7 <i>76.9</i>	3.7	0.0261	1.9	0.0787
PGP-CHP (g)	8.4 <i>2.3</i>	9.5 <i>3.1</i>	10.7 <i>3.3</i>	11.7 <i>3.3</i>	9.9 <i>3.2</i>	10.9 <i>3.4</i>	10.6 <i>3.2</i>	10.4 <i>3.4</i>	4.5	0.0116	2.3	0.0313
NG_E-CHP	44.1 <i>10.7</i>	43.7 <i>8.6</i>	51.5 <i>6.6</i>	49.7 <i>7.7</i>	46.2 <i>9.3</i>	49.2 <i>8.4</i>	46.9 <i>6.3</i>	43.0 <i>9.0</i>	17.1	<0.0001	2.7	0.0145
P1000G-CHP (g)	45.1 <i>6.5</i>	42.8 <i>5.3</i>	42.7 <i>4.3</i>	48.5 <i>5.5</i>	42.2 <i>4.8</i>	42.9 <i>4.7</i>	43.6 <i>4.4</i>	44.2 <i>5.2</i>	21.7	<0.0001	7.8	<0.0001

Rappel des abréviations : Hd=date d'épiaison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante, NG\_E=nombre de grains par épi, P1000G=poids de 1000 grains ; CHP indique que les caractères considérés ici sont ceux mesurés au champ. Dans le modèle d'analyse de la variance : *Pop*=effet population et *Gene(Pop)*= effet génération dans la population.

### IV-3.2 Evolution de la variabilité intra population pour les caractères mesurés au champ

Les différentes estimations de variabilité génétique intra population à chaque génération sont présentés dans le Tableau IV-4. Les héritabilités estimées sont élevées et significativement différentes de zéro pour la plupart des caractères. Certaines composantes de la productivité présentent des héritabilités non significativement différentes de zéro en génération 7 de la population du Moulon et en génération 2 de la population de Toulouse. Ces faibles héritabilités sont liées à de forts coefficients de variations environnementales couplés à de plus faibles coefficients de variations génétiques que dans les autres populations. Le coefficient de variation génétique observé pour la hauteur diminue sensiblement au cours des générations dans les populations de Vervins et Le Moulon tandis qu'elle augmente dans la population de Toulouse. En revanche aucune tendance nette d'évolution des variances ou de l'héritabilité ne se dégage pour les autres caractères. Pour la plupart des caractères, les coefficients de variation génétiques sont maintenus au cours du temps et ne présentent pas de baisse importante par rapport à la population initiale.

Tableau IV-4. Variances génétiques (VG) et environnementales (VE), héritabilité (H<sup>2</sup>) et coefficients de variation (CV-G) et environnementaux (CV-E) estimés dans chacune des populations étudiées

Population	P. initiale			Vervins			Le Moulon			Toulouse											
	G0	G2	G7	G2	G7	G12	G2	G7	G12	G2	G7	G12									
HdGHP	VG <i>s.d.</i>	2869.3	714.3	2793.3	554.6	3701.6	733.0	2177.6	416.9	1881.3	372.3	1384.7	282.0	1815.6	390.1	3268.1	677.8	1932.4	393.1	2441.0	491.9
	VE <i>s.d.</i>	225.4	38.9	263.7	36.2	272.7	39.4	172.9	24.3	243.0	34.5	230.1	32.8	436.9	64.8	327.5	48.8	293.3	41.1	180.5	26.5
	H <sup>2</sup>	<b>0.93</b> ***		<b>0.93</b> ***		<b>0.93</b> ***		<b>0.93</b> ***		<b>0.81</b> ***		<b>0.86</b> ***		<b>0.81</b> ***		<b>0.91</b> ***		<b>0.87</b> ***		<b>0.93</b> ***	
HTGHP	CV-G (%)	4.4		4.8		3.6		3.6		3.5		3.0		3.5		4.7		3.6		4.1	
	CV-E (%)	1.2		1.3		1.0		1.1		1.3		1.2		1.7		1.5		1.4		1.1	
	VG <i>s.d.</i>	132.2	34.4	190.2	38.3	214.0	42.9	126.3	25.2	216.2	43.4	126.4	25.4	82.1	17.0	90.6	27.1	155.0	30.8	173.1	35.2
VE <i>s.d.</i>	28.6	4.8	25.6	3.4	19.9	2.7	27.3	3.6	28.4	3.8	15.7	2.1	16.1	2.2	125.8	17.6	15.5	2.1	19.6	2.7	
H <sup>2</sup>	<b>0.82</b> ***		<b>0.88</b> ***		<b>0.92</b> ***		<b>0.82</b> ***		<b>0.88</b> ***		<b>0.89</b> ***		<b>0.84</b> ***		<b>0.42</b> ***		<b>0.91</b> ***		<b>0.90</b> ***		
NEGHP	CV-G (%)	12.9		14.0		9.5		9.5		16.0		10.4		8.7		10.2		12.2		13.0	
	CV-E (%)	6.0		5.1		4.4		4.4		5.8		3.7		3.9		12.0		3.9		4.4	
	VG <i>s.d.</i>	0.8	0.4	1.1	0.4	0.5	0.3	0.6	0.3	1.2	0.4	0.8	0.4	1.0	0.4	0.7	0.4	0.6	0.3	0.6	0.4
VE <i>s.d.</i>	2.4	0.4	2.4	0.3	2.6	0.4	2.2	0.3	2.4	0.3	3.6	0.5	3.1	0.4	3.5	0.5	2.3	0.3	3.9	0.5	
H <sup>2</sup>	<b>0.25</b> **		<b>0.32</b> ***		<b>0.16</b> *		<b>0.20</b> *		<b>0.33</b> ***		<b>0.19</b> (ns)		<b>0.25</b> *		<b>0.16</b> (ns)		<b>0.20</b> *		<b>0.13</b> (ns)		
NGPCHP	CV-G (%)	17.5		21.4		15.6		15.6		21.2		16.8		19.9		15.9		16.1		14.4	
	CV-E (%)	30.0		31.3		30.7		30.7		30.0		35.1		34.0		35.9		31.8		36.5	
	VG <i>s.d.</i>	2128.7	1007.1	2975.6	1058.1	2438.4	891.0	2263.1	908.1	3074.4	1000.7	1278.6	939.5	3673.8	1490.6	1231.8	1121.2	2290.3	820.0	2533.4	1249.0
VE <i>s.d.</i>	5473.2	925.1	6852.7	924.0	5825.7	792.8	6976.0	912.1	5831.8	786.4	9285.7	1252.1	10701.4	1461.6	11186.9	1566.5	5350.4	721.4	10137.7	1392.5	
H <sup>2</sup>	<b>0.28</b> **		<b>0.30</b> ***		<b>0.30</b> **		<b>0.24</b> *		<b>0.35</b> ***		<b>0.12</b> (ns)		<b>0.26</b> *		<b>0.10</b> (ns)		<b>0.30</b>		<b>0.20</b> *		
PGPCHP	CV-G (%)	20.8		21.8		19.8		19.8		23.5		13.4		23.7		14.4		22.4		21.4	
	CV-E (%)	33.4		33.1		34.8		34.8		32.4		36.1		40.5		43.4		34.2		42.7	
	VG <i>s.d.</i>	5.8	2.4	6.8	2.2	8.1	2.6	4.6	2.1	6.4	2.0	3.0	2.1	5.5	2.4	3.4	2.2	4.4	1.6	4.8	2.4
VE <i>s.d.</i>	10.8	1.8	12.4	1.7	14.4	2.0	18.0	2.4	10.7	1.4	19.8	2.7	18.9	2.6	20.1	2.8	10.5	1.4	19.5	2.7	
H <sup>2</sup>	<b>0.35</b> **		<b>0.36</b> ***		<b>0.36</b> ***		<b>0.20</b> *		<b>0.38</b> ***		<b>0.13</b> (ns)		<b>0.22</b> *		<b>0.14</b> (ns)		<b>0.29</b> **		<b>0.20</b> *		
NG_EGHP	CV-G (%)	25.3		24.4		18.4		18.4		25.6		15.2		21.5		17.3		22.3		21.1	
	CV-E (%)	34.5		32.9		36.5		36.5		33.0		38.6		40.1		42.2		34.7		42.4	
	VG <i>s.d.</i>	60.1	18.2	17.2	9.1	43.8	12.1	38.5	11.3	69.3	16.8	43.9	12.3	55.7	13.7	20.5	8.5	53.6	13.9	64.0	16.0
VE <i>s.d.</i>	43.1	7.3	78.8	10.6	52.1	7.1	61.4	8.0	51.1	6.9	55.6	7.5	39.2	5.4	59.2	8.3	53.1	7.2	50.0	6.9	
H <sup>2</sup>	<b>0.58</b> ***		<b>0.18</b> *		<b>0.46</b> ***		<b>0.39</b> ***		<b>0.58</b> ***		<b>0.44</b> ***		<b>0.59</b> ***		<b>0.26</b> *		<b>0.50</b> ***		<b>0.56</b> ***		
P1000GHP	CV-G (%)	17.8		8.0		12.5		12.5		18.0		13.3		15.3		9.7		16.4		18.6	
	CV-E (%)	15.0		17.2		15.8		15.8		15.5		15.0		12.8		16.4		16.3		16.4	
	VG <i>s.d.</i>	27.1	6.9	16.6	3.5	25.4	5.5	28.1	5.7	20.5	4.4	18.8	4.1	19.2	4.2	13.7	3.9	10.0	2.6	25.6	5.3
VE <i>s.d.</i>	4.3	0.7	5.0	0.7	8.5	1.2	7.4	1.0	7.4	1.0	7.6	1.0	7.5	1.0	10.6	1.5	10.1	1.4	4.0	0.5	
H <sup>2</sup>	<b>0.86</b> ***		<b>0.77</b> ***		<b>0.75</b> ***		<b>0.79</b> ***		<b>0.73</b> ***		<b>0.71</b> ***		<b>0.72</b> ***		<b>0.60</b> ***		<b>0.50</b> ***		<b>0.87</b> ***		
P1000GHP	CV-G (%)	12.2		9.5		10.9		10.9		10.7		10.1		10.2		9.1		7.2		11.4	
	CV-E (%)	4.8		5.2		5.6		5.6		6.5		6.4		6.4		7.5		7.3		4.5	

Rappel des abréviations: Hd=date d'épison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=pois de grains par plante, NG\_E=nombre de grains par épi, P1000G=pois de 1000 grains.

GHP indique que les caractères considérés ici sont ceux mesurés au champ.

Pour H<sup>2</sup>: (ns)= non significatif, \* = P<0.05, \*\* = P<0.01, \*\*\*= P<0.001

### IV-3.3 Détection de la sélection

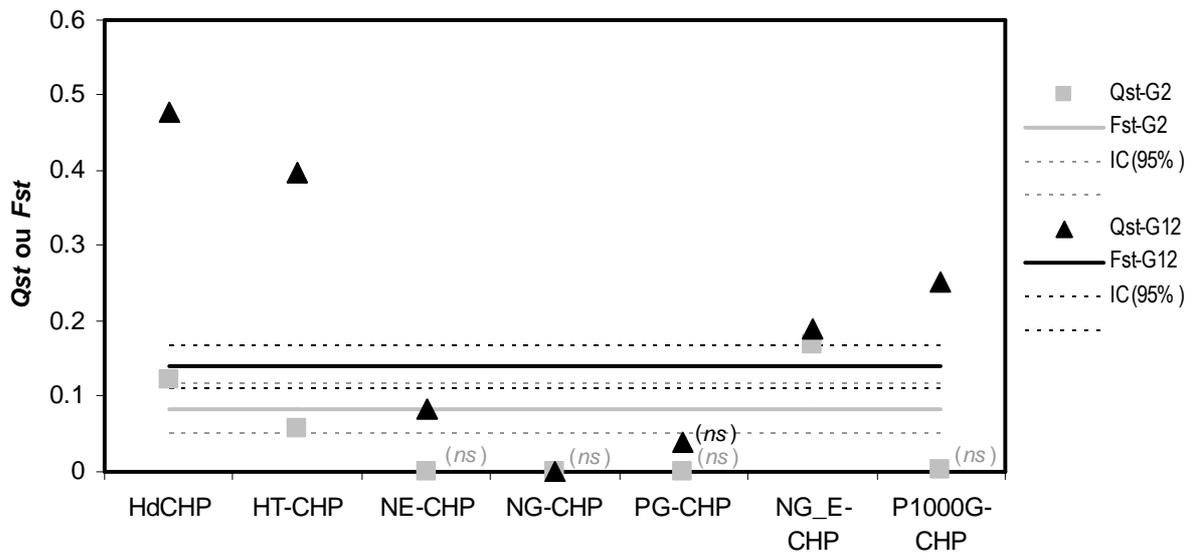


Figure IV-2 : Estimation de la différenciation entre les 3 populations de gestion dynamique pour la précocité et les composantes de la valeur sélective des populations ( $Q_{ST}$ ) en génération 2 (en gris) et en génération 12 (en noir), et confrontation avec les valeurs de différenciation estimées pour les marqueurs microsatellites ( $F_{ST}$ ) aux 2 générations considérées. Toutes les valeurs de  $Q_{ST}$  sont significatives ( $P < 0,01$ ) sauf mention contraire (ns). Les lignes en pointillées donnent les intervalles de confiance à 95% estimés par bootstrap sur les locus (cf. chapitre III).

En génération 2, les niveaux de différenciation ( $Q_{ST}$ ) estimés pour l'ensemble des caractères évalués au champ sont proches de la différenciation estimée à partir de la variation inter populations des 21 locus microsatellites répartis sur l'ensemble du génome (Figure IV-2). En génération 12, les valeurs de différenciation estimées pour la précocité de floraison et la hauteur sont largement supérieures au  $F_{ST}$  estimé à partir des marqueurs microsatellites, avec une valeur 3,5 fois supérieure pour la précocité de floraison et près de 3 fois supérieure pour la hauteur. Certaines valeurs de différenciation restent en revanche très faibles et non significativement différentes de zéro pour des composantes de la valeur sélective (Poids de grains et nombre de grains par plante).

#### **IV-3.4 Corrélations entre précocité et composantes de la valeur sélective : estimation des gradients de sélection.**

Afin d'estimer la force de sélection agissant directement sur la précocité de floraison et ayant pu conduire à un tel niveau de différenciation entre les populations, les gradients de sélection tenant compte de la sélection d'autres caractères potentiellement corrélés ont été estimés à partir de la réponse à la sélection observée entre 2 générations et des corrélations génétiques entre caractères estimées en génération 2 dans chacune des populations.

Pour l'ensemble des populations, les gradients de sélection traduisant la force de sélection agissant directement sur les caractères à chaque génération sont faibles (Tableau IV-4), mais sont du même ordre de grandeur que ceux estimés dans la thèse de J. David entre la génération initiale G0 et la génération 4 de la population PB de gestion dynamique. La population de Vervins est celle qui a subi les plus forts gradients de sélection. La hauteur, le poids de 1000 grains et dans une moindre mesure la précocité de floraison évoluent significativement entre la génération 2 et la génération 12, mais seule la hauteur semble avoir été sélectionnée de façon directe, tandis que le poids de 1000 grains a évolué corrélativement avec la hauteur, les plantes plus hautes produisant des grains plus remplis. Le nombre de grains par épi présente, en valeur absolue, le plus fort gradient mais pourtant ce caractère évolue peu entre la génération 2 et la génération 12, suggérant des pressions de sélection antagonistes liées à la corrélation génétique négative entre ce caractère et le poids de 1000 grains. Pour la population du Moulon, seule la hauteur évolue de façon significative en relation avec un gradient de sélection plus fort que pour les autres caractères. Dans la population de Toulouse, la hauteur et le nombre de grains par épi présentent les plus forts gradients de sélection conduisant à une évolution vers des plantes plus hautes et produisant moins de grains par épi. Mais cette faible productivité par épi est compensée par une augmentation du nombre d'épis par plante. Quelle que soit la population considérée, la précocité de floraison présente des gradients de sélection faibles ce qui suggère que ce caractère n'est pas sélectionné de façon directe.

Tableau IV-4 : Réponse à la sélection ( $\bar{z}_{12} - \bar{z}_2$ ) observée entre la génération 2 et la génération 12 de chaque population et gradient de sélection stabilisante  $\beta$ .

	$\bar{z}_{12} - \bar{z}_2$	$\beta$	Matrices des corrélations génétiques en G2				
Vervins			HdCHP	HT-CHP	NE-CHP	NG_E-CHP	P1000G-CHP
HdCHP	0.533	0.025	HdCHP				
HT-CHP	1.362	0.046	HT-CHP	0.372			
NE-CHP	-0.043	0.029	NE-CHP	-0.197	-0.074		
NG_E-CHP	-0.196	-0.072	NG_E-CHP	0.011	-0.110	0.053	
P1000G-CHP	1.258	0.022	P1000G-CHP	0.266	0.568	0.022	-0.328
Le Moulon			HdCHP	HT-CHP	NE-CHP	NG_E-CHP	P1000G-CHP
HdCHP	-0.150	-0.013	HdCHP				
HT-CHP	0.806	0.031	HT-CHP	0.083			
NE-CHP	-0.004	0.001	NE-CHP	-0.278	-0.033		
NG_E-CHP	0.245	0.012	NG_E-CHP	0.406	0.212	-0.108	
P1000G-CHP	0.124	-0.006	P1000G-CHP	-0.263	0.170	0.020	-0.164
Toulouse			HdCHP	HT-CHP	NE-CHP	NG_E-CHP	P1000G-CHP
HdCHP	-0.339	-0.016	HdCHP				
HT-CHP	0.545	0.055	HT-CHP	0.281			
NE-CHP	0.116	-0.014	NE-CHP	-0.340	-0.263		
NG_E-CHP	-0.430	-0.059	NG_E-CHP	0.275	-0.031	-0.174	
P1000G-CHP	0.129	-0.005	P1000G-CHP	-0.381	-0.126	0.309	-0.311

Rappel des abréviations : Hd=date d'épiaison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante, NG\_E=nombre de grains par épi, P1000G=poids de 1000 grains. CHP indique que les caractères considérés ici sont ceux mesurés au champ.

### IV-3.5 Estimation des gradients de sélection attendus en conditions différentes pour la durée de vernalisation.

Les différentielles et les gradients de sélection attendus en fonction de la longueur de la période de vernalisation ont été estimés en génération 2 dans chacune des populations séparément. Comme les résultats sont très concordants entre les 3 populations considérées et pour augmenter le jeu de données, les différentielles et gradients de sélection présentés ici sont les estimations faites en génération 2 pour ces 3 populations confondues.

Figure IV-3 : Différentielles (à gauche) et gradient de sélection (à droite) estimées à partir de la génération 2 toutes populations confondues dans des environnements différents pour la durée de la période de croissance (CHP, 8S) et la durée de la période de vernalisation (8S, 4S). Pour les gradients, les bars d'erreurs correspondent à l'écart-type du coefficient de régression.

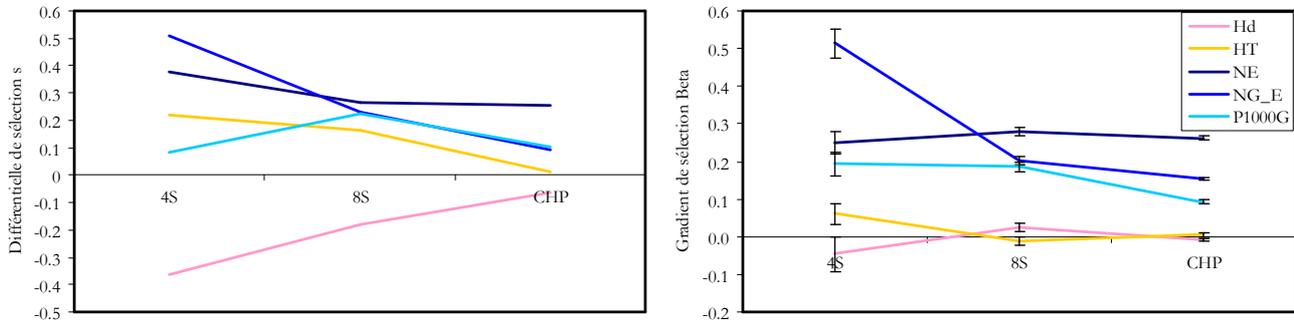
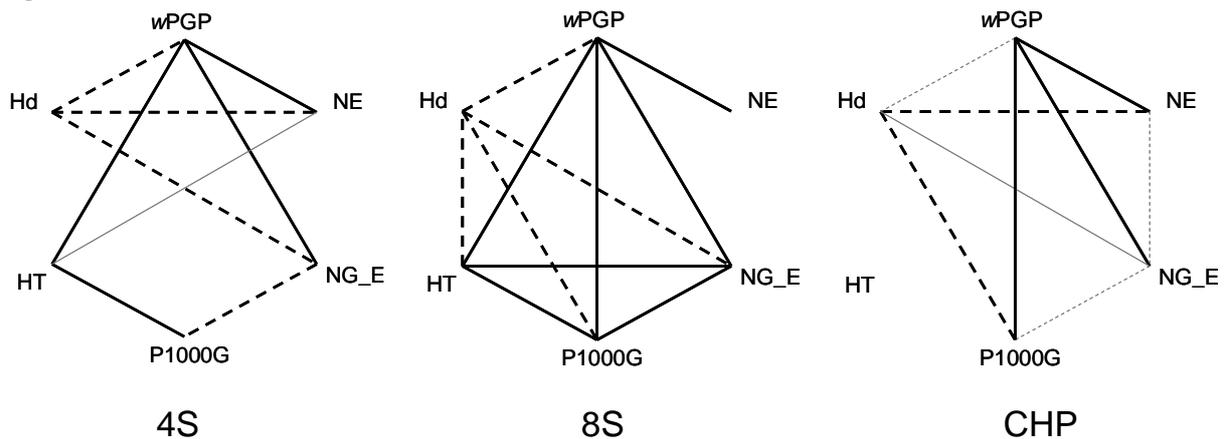


Figure IV-4 : Corrélations phénotypiques significatives entre caractères en génération 2 dans les différentes conditions testées. Les traits pleins indiquent des corrélations positives, les traits en pointillés indiquent les corrélations négatives. Les traits gris indiquent que la corrélation est significative au seuil de 5%, les traits noirs indiquent que la corrélation est significative au seuil de 1%.



En conditions normales de champ, seules les composantes de la valeur sélective sont à la fois directement et indirectement sélectionnées (corrélations et coefficients de régression significatifs entre ces caractères et la valeur sélective relative, Figures IV-5 et IV-6). La précocité de floraison est sélectionnée dans une moindre mesure de façon indirecte du fait de la corrélation négative avec le nombre d'épis et le poids de 1000 grains qui sont tous les deux sélectionnés. Mais la sélection indirecte pour des génotypes plus précoces est contrecarrée par la sélection du nombre de grains par épi positivement corrélé au caractère de précocité (Figure IV-4). En conditions 8S, les différentielles de sélection (*i.e.* la contribution indirecte des caractères phénotypiques au succès reproducteur) sont significatives pour la précocité et la

hauteur allant dans le sens d'une sélection de génotypes plus précoces, plus hauts et produisant plus d'épis. La précocité est sélectionnée de façon indirecte à travers les corrélations négatives avec les autres caractères. En conditions de vernalisation partielle (4S), les plantes précoces plus hautes et produisant beaucoup de grains sont sélectionnées. En revanche, les gradients de sélection (*i.e.* la contribution directe de ces caractères phénotypiques au succès reproducteur) sont non significatifs pour la hauteur et la précocité, quelque soit la durée de la période de vernalisation.

### **IV-3.6. Evolution des composantes de la précocité de floraison**

Dans cette partie, les mesures de précocité dans des environnements variables de vernalisation (conditions NV, 4S et 8S) et de longueur de jour (condition JC) sont comparées pour estimer l'évolution de la sensibilité aux facteurs environnementaux et de la précocité intrinsèque dans les différentes populations.

#### **IV-3.6.1 Evolution des composantes de la précocité dans les populations**

L'évolution de la sensibilité à la longueur du jour et à la vernalisation des différentes populations peut être appréhendée en évaluant la précocité en conditions contrôlées.

Ainsi, l'évaluation de la précocité en conditions de jours courts (HdJC) permet de distinguer les génotypes sensibles et les insensibles à la longueur du jour (Figure IV-5). La précocité de floraison ainsi évaluée présente une distribution asymétrique liée à une part importante de génotypes n'ayant pas épié avant la fin de l'expérimentation ( $HdJC=3000^{\circ}J$ ) qui peuvent être considérés comme fortement sensibles à la photopériode (Figure IV-5). Pour les génotypes ayant fleuri durant le temps de l'expérimentation, l'épiaison s'est faite par paliers avec une première série d'épiaisons entre 1200 et 1700 $^{\circ}J$  pour les génotypes insensibles à la photopériode, une deuxième série d'épiaisons entre 1700 et 2300 $^{\circ}J$  et une troisième série entre 2300 et 2800 $^{\circ}J$ . En cours d'expérimentation, certains génotypes sensibles à la photopériode ( $HdJC>1700^{\circ}J$ ) présentent des anomalies de développements importantes (Figure IV-6).

Figure IV-5 : Evolution des différentes mesures de la précocité de floraison évaluée dans les populations aux cours des générations. A. Distribution du caractère toutes populations confondues ; B. Distribution du caractère dans les 3 populations étudiées et évolution au cours des générations ; C. Evolution des moyennes par population au cours de générations. Les barres d'erreurs correspondent aux intervalles de confiance à 95% estimés sur la moyenne.

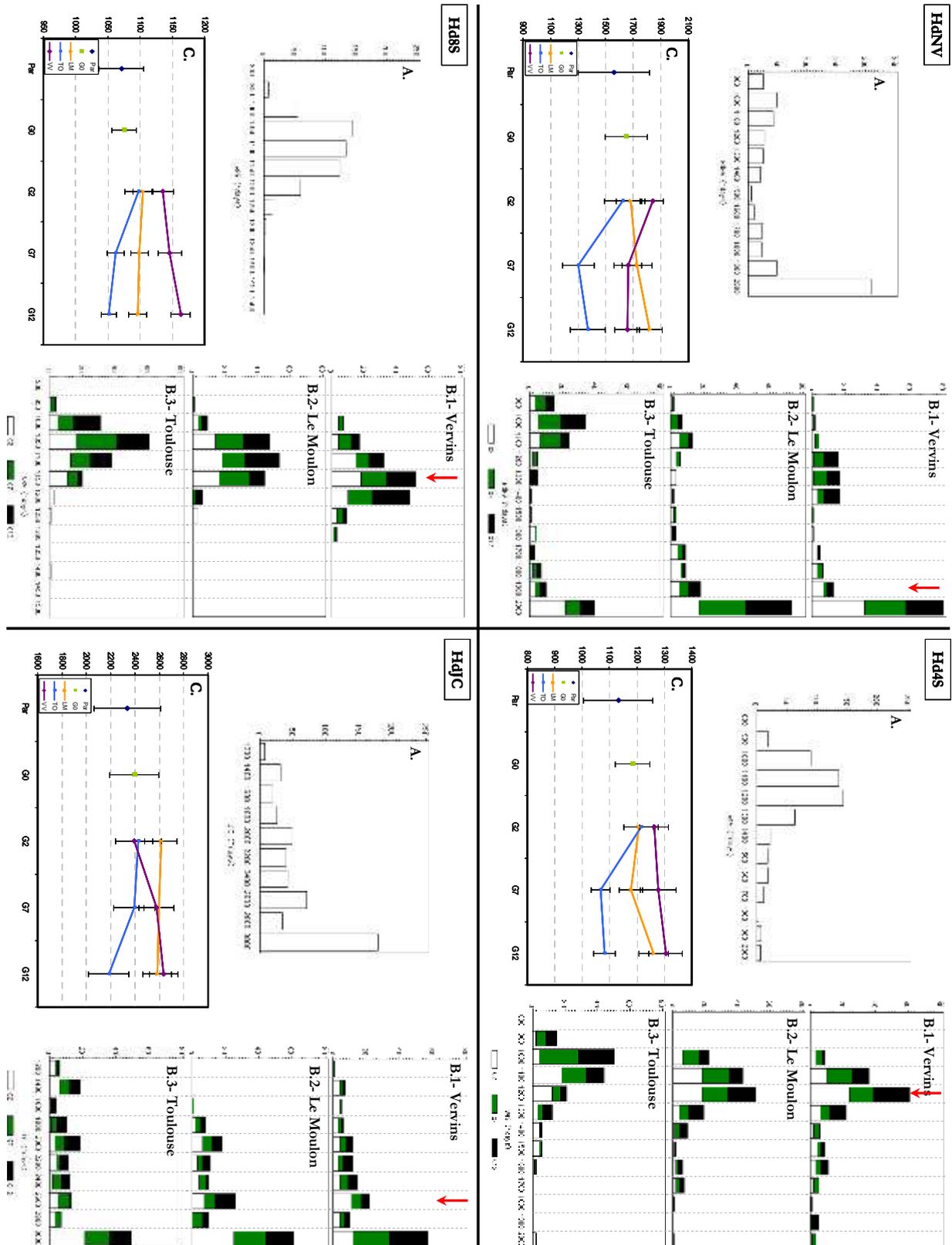


Figure IV-6 : Exemple d'anomalies de développement observées parmi les génotypes sensibles en condition de jours courts. Photo de gauche : l'épi en formation reste bloqué dans la gaine ; photo de droite : la plante présente un renflement de la gaine caractéristique de la phase de pré-épiaison mais aucun épi ne sort de la gaine et une nouvelle feuille est initiée.



Dans la population de Toulouse, la fréquence des génotypes insensibles ( $HdJC < 1700^\circ J$ ) à la longueur du jour ( $f = 0,22$  en G12) est du même ordre que ce qui est observé dans la population parentale ( $f = 0,25$ ) et la population initiale ( $f = 0,26$ ), tandis que dans les populations du Nord, la proportion de génotypes insensibles diminue fortement jusqu'à 0,02 dans la population de Vervins en G12 et jusqu'à disparition complète dans la population du Moulon (Figure IV-8).

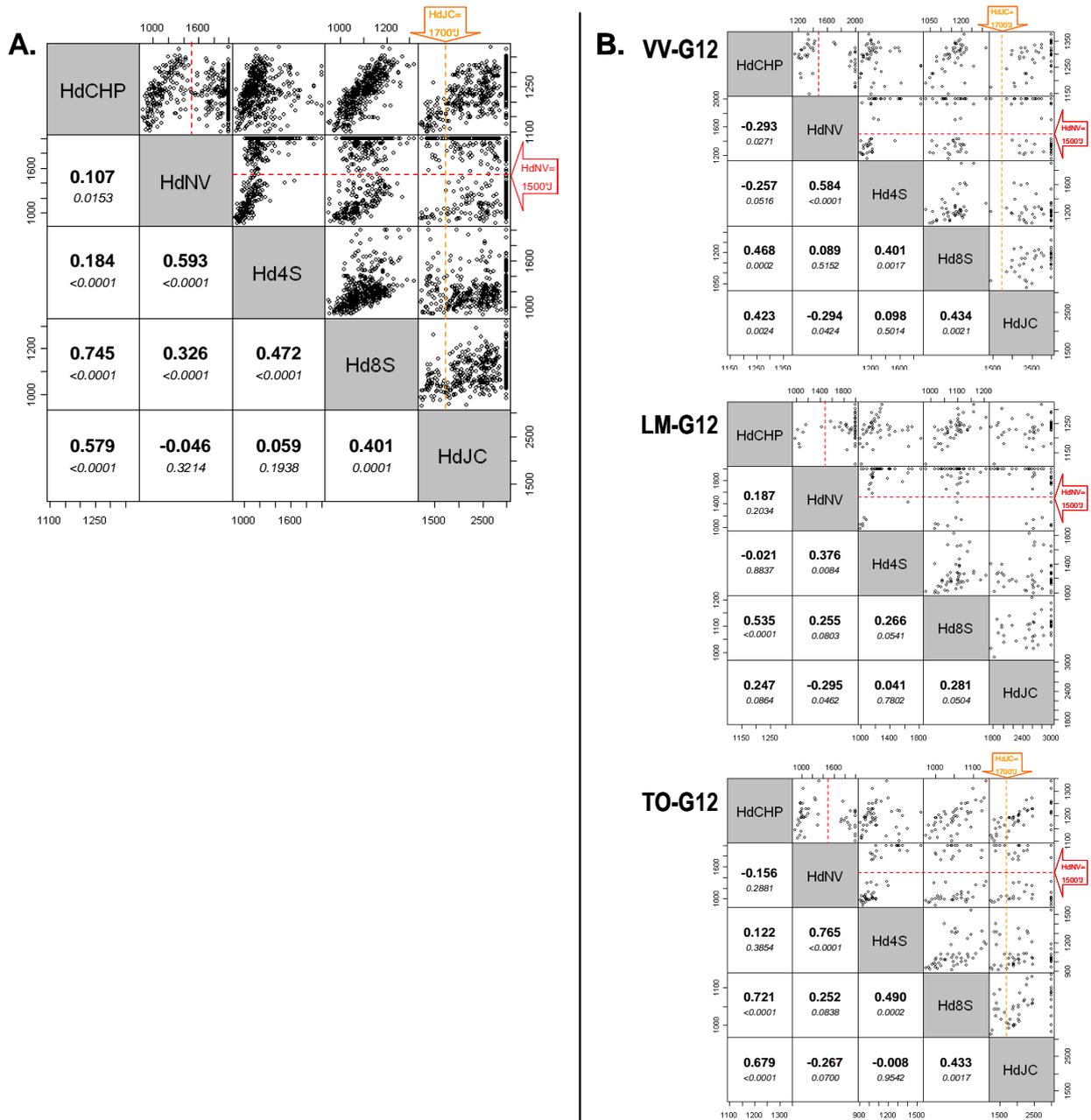
L'évaluation de la précocité en l'absence de vernalisation ( $HdNV$ , Figure IV-5) permet de distinguer les génotypes sensibles à la vernalisation (présentant un délai de floraison ou ne fleurissant pas en l'absence de vernalisation). La distribution du caractère de précocité de floraison évaluée en l'absence de vernalisation ( $HdNV$ ) est de type bimodale avec une catégorie de génotypes épiant avant  $1500^\circ J$  et une autre épiant après les  $1500^\circ J$  ou n'ayant pas épié avant la fin de l'expérimentation ( $HdNV = 2000^\circ J$ ). La valeur seuil de  $1500^\circ J$  a été choisie pour différencier les génotypes de type "Printemps" (P, épiant avant  $1500^\circ J$ ) des génotypes de type "Hiver" (H) qui est une distinction courante chez le blé pour caractériser les blés nécessitant ou non une période de froid pour initier la floraison. La distribution des types H et P évolue au cours du temps avec une augmentation significative des génotypes P au cours des générations dans la population de Toulouse et une augmentation significative des génotypes H dans la population du Moulon tandis que dans la population de Vervins, la répartition H/P est plus variable dans le temps avec dans un premier temps une augmentation significative de la

fréquence des types H en G2 suivie par la suite d'une augmentation de la fréquence des type P en G7 et G12 (cf. Rhoné *et al.* 2008, article présenté en chapitre V). Notons que les génotypes de type P dans la population de Vervins tendent à être plus tardifs que les génotypes P des autres populations. L'expérimentation visant à évaluer la précocité des génotypes en conditions partielles de vernalisation (Hd4S, Figure IV-5) montre que parmi les génotypes sensibles au froid certains ne sont pas satisfaits avec 4 semaines de vernalisation et présentent un délais de floraison en conditions de vernalisation partielle (épiaison entre 1400 et 1800°J) ou n'épient pas avant la fin de l'expérimentation (Hd4S=2000°J). En condition partielle de vernalisation, la population de Toulouse se démarque nettement des deux autres populations en fleurissant de façon plus précoce dès la génération 7. Parmi les composantes de la précocité, seule la précocité intrinsèque (Hd8S) présente une évolution divergente bien nette entre les 3 populations suivant la même évolution que *HdCHP* avec une épiaison de plus en plus tardive pour la population de Vervins et une épiaison plus précoce pour la population de Toulouse (Figure IV-5).

### **IV-3.6.2 Evolution des corrélations entre mesures de précocité**

Les corrélations phénotypiques entre les différentes mesures de précocité sont positives et fortement significatives lorsque l'on considère toutes les populations ensemble, sauf entre la précocité évaluée en conditions de jours courts (HdJC) et la précocité évaluée en absence de vernalisation (HdNV) ou en vernalisation partielle (Hd4S ; Figure IV-7 A). La plus forte valeur de corrélation est trouvée entre la précocité au champ et la précocité intrinsèque pour lesquelles les conditions en vernalisation et en photopériode sont satisfaites. La précocité observée au champ est également fortement corrélée à la précocité observée en conditions de jours courts, les génotypes les plus précoces au champ étant tous des individus identifiés comme insensibles à la longueur du jour (HdJC<1700°J), tandis que les génotypes sensibles fleurissent tous au champ avec un décalage de 50°J ( $\approx 2$  jours) par rapport aux insensibles (Figure IV-7 A). Les corrélations observées en l'absence de vernalisation sont fortement dépendantes du caractère Hiver ou Printemps des génotypes (Figure IV-7). En particulier, une forte corrélation est observée entre HdCHP et HdNV pour les génotypes de type P ( $R=0,63$  ;  $P<0,0001$ ) alors que la corrélation est nulle pour les génotypes de type H ( $R=0,03$  ;  $P=0,6$ ). Les génotypes P étant insensible à la vernalisation, la corrélation observée avec l'épiaison au champ pour ces génotypes est donc le fait de la précocité intrinsèque.

Figure IV-7 : Matrices des corrélations phénotypiques, toutes populations confondues (A) et dans les différentes populations étudiées en G12 (B) entre les mesures d'épaisseur des populations de gestion dynamique dans différentes conditions de vernalisation et de longueur du jour. Les distributions conjointes des différentes mesures sont représentées au dessus de la diagonale. La ligne pointillée rouge indique la valeur de HdNV séparant les types Hiver des types Printemps, la ligne pointillée orange indique la valeur de HdJC séparant les génotypes sensibles à la longueur du jour des génotypes insensibles. Les valeurs de corrélation de Pearson sont indiquées en dessous de la diagonale, les valeurs en italique correspondent aux probabilités associées aux valeurs de corrélations ( $H_0 : r = 0$ ).

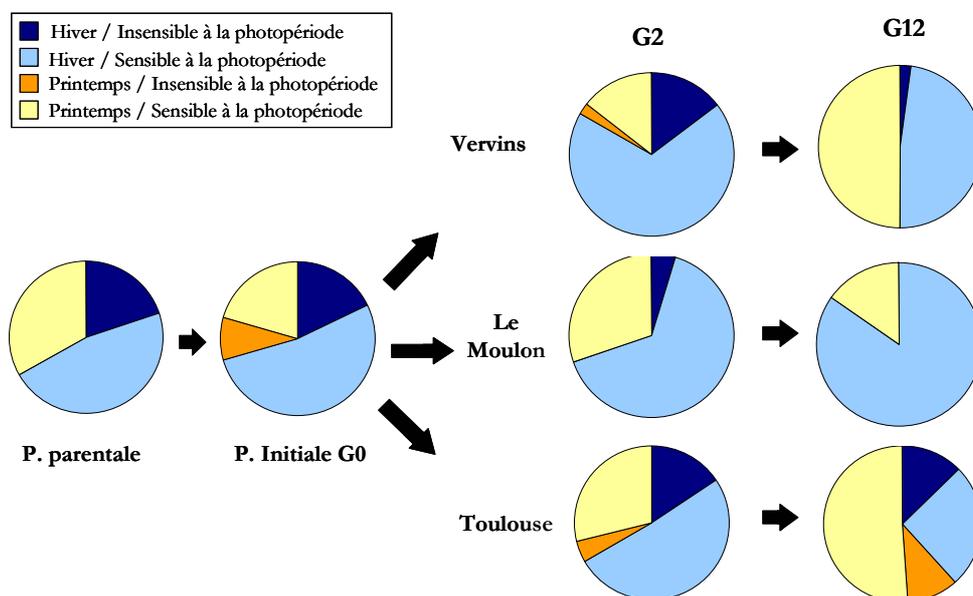


Certains génotypes H fleurissent de façon précoce au champ alors que les conditions de vernalisation sont satisfaites ce qui traduit que, pour ces génotypes, l'absence de vernalisation

induit un décalage important de la floraison. En revanche pour les génotypes de type H épiant de façon tardive au champ, la part de besoin en vernalisation est difficile à dissocier de la part de précocité intrinsèque dans l'établissement de la date de floraison.

Dans la population de Vervins en génération 12, les génotypes de type P sont ceux qui fleurissent le plus tôt en condition de vernalisation incomplète (Hd4S) mais qui tendent à être les plus tardifs au champ, ce qui se traduit par une corrélation négative entre HdNV et HdCHP (Figure IV-7 B). La population est composée pour moitié de génotype P et pour moitié de génotypes H dans la dernière génération observée (Figure IV-8). Dans la population de Toulouse, une part des génotypes de type H sensibles au besoin en vernalisation épie précocement au champ indiquant que, pour ces génotypes, la sensibilité à la vernalisation est un signal majeur. En parallèle, les génotypes précoces au champ sont tous insensibles à la longueur du jour. Ces deux observations suggèrent que la population de Toulouse en génération 12 est composée essentiellement de génotypes sensibles à l'un ou l'autre signal environnemental (longueur du jour ou vernalisation). Cependant une faible part de génotypes insensibles aux deux signaux environnementaux à la fois, déjà observés dans la population initiale, subsistent dans la population de Toulouse jusqu'en génération 12 (Figure IV-8). Ces génotypes sont également ceux qui épiant le plus précocement au champ. En revanche, dans la population du Moulon, il n'existe plus aucun génotype insensible à la longueur du jour et une majorité des génotypes est sensible à la vernalisation, cumulant ainsi la sensibilité aux deux facteurs environnementaux.

Figure IV-8 : Distribution des types hiver / printemps, sensibles / insensibles à la photopériode dans les 3 populations des gestion dynamique.



## IV-4 Discussion

### IV-4.1 Evolution de la précocité: mise en évidence de la sélection divergente

L'évaluation de la précocité au champ pour l'ensemble des populations étudiées menée sur le site du Moulon a mis en évidence que les populations évoluent de façon divergente conduisant à l'établissement d'un gradient Nord-Sud de précocité, les populations du Nord fleurissant plus tardivement que la populations du Sud. Cette évolution se traduit par un niveau de différenciation croissant aux cours des générations. Initialement proche du niveau de différenciation estimé à partir de 21 locus microsatellites répartis sur l'ensemble du génome ( $F_{ST}$ ), le niveau de différenciation pour la précocité ( $Q_{ST}$ ) augmente plus vite que le  $F_{ST}$  entre la génération 2 et la génération 12. En G12, le  $Q_{ST}$ , 4 fois supérieur au  $F_{ST}$ , peut être vu comme une marque de sélection divergente agissant sur les populations. En effet, la différenciation des locus microsatellites est interprétée comme une estimation du niveau de différenciation subit par les populations du simple fait de la dérive. Ainsi, un caractère quantitatif à base génétique purement additive évoluant dans différentes populations sous le seul effet de la dérive aura un niveau de différenciation du même ordre que celui observé pour les marqueurs (Whitlock 1999). A partir de ce postulat, la comparaison  $Q_{ST} / F_{ST}$  peut conduire à 3 types de conclusions :

(i)  $Q_{ST} > F_{ST}$  : la part de variance entre populations dans la variance totale observée pour le caractère considéré est supérieure à celle observée pour les marqueurs supposés neutres. Ceci traduit le fait que des phénotypes différents ont été sélectionnés dans les différents environnements et est interprété comme une preuve de sélection divergente agissant sur le caractère étudié dans les populations.

(ii)  $Q_{ST} < F_{ST}$  : à l'inverse du cas précédent, les mêmes phénotypes sont favorisés dans des environnements différents, traduisant l'action d'une sélection homogénéisante sur le caractère.

(iii)  $Q_{ST} = F_{ST}$  : un niveau de différenciation équivalent entre caractères et marqueurs neutres traduit simplement le fait que la différenciation observée pour le caractère a pu être le fait de la dérive seule. Dans cette situation, on ne peut faire la distinction entre les effets de la dérive et les effets de la sélection agissant potentiellement sur le caractère (Merillä et Crnokrak 2001).

La comparaison  $Q_{ST} / F_{ST}$  est très couramment utilisée pour distinguer la divergence entre populations relative aux effets de la dérive de celle attribuable aux effets de la sélection naturelle et mettre ainsi en évidence l'adaptation locale de population. Une méta-analyse récente (Leinonen *et al.* 2008) fait état de 62 publications recensées dans la bibliographie (dont 34 ces 4 dernières années) pour des organismes variés tels qu'*Arabidopsis thaliana* (Porcher *et al.* 2006 ; Le Corre 2005), la grenouille commune (Palo *et al.* 2003) ou encore le pin sylvestre (Waldmann *et al.* 2005). Le succès de cette approche pour détecter la sélection provient du fait que c'est un outil simple à mettre en œuvre et qui ne repose pas sur un modèle évolutif particulier (Whitlock 1999). Cependant, l'utilisation de cette approche a également fait l'objet de nombreux débats liés aux biais d'estimation de la différenciation des marqueurs neutres ou des caractères quantitatifs et qui incite à considérer cette approche avec précautions lors de son utilisation sur nos données :

→ Biais pouvant affecter l'estimation du  $F_{ST}$  :

La comparaison  $Q_{ST} / F_{ST}$  pour mettre en évidence de la sélection repose sur l'hypothèse fondamentale que les locus considérés pour estimer le  $F_{ST}$  évoluent de façon neutre et ne sont pas soumis à sélection. L'utilisation de marqueurs liés à des locus soumis à sélection peut conduire à un biais d'estimation (sur ou sous-estimation selon que le locus est soumis à de la sélection divergente ou homogénéisante), le niveau de différenciation ainsi estimé ne reflète donc pas le niveau de divergence lié à la dérive pure. D'autres biais peuvent apparaître en liaison avec l'utilisation de marqueurs microsatellites pour estimer le  $F_{ST}$ . D'une part, la forte diversité généralement observée pour ce type de marqueurs est susceptible d'induire une variance intra population forte, d'autre part le risque d'homoplasie en relation avec le mode de mutation des marqueurs microsatellites tend à sous-estimer la variance inter populations. Ces deux processus peuvent conduire à une sous-estimation de la différenciation entre populations. Dans le travail présenté ici, 21 locus microsatellites répartis sur l'ensemble du génome sont utilisés pour estimer la différenciation « neutre » ce qui correspond au double du nombre moyen de marqueurs utilisés dans les études listées dans la méta-analyse citée

précédemment (nombre moyen de locus = 9, SE = 8, sans considérer 4 études particulières utilisant plus de 100 locus ; Leinonen 2008). L'utilisation d'un grand nombre de locus permet de diluer les effets spécifiques d'un seul locus se comportant de façon non neutre et on peut penser que la sélection indirecte sur un locus dans un sens ou dans un autre, du fait de sa liaison à un gène sélectionné, affectera peu l'estimation globale moyenne. De plus l'utilisation de marqueurs microsatellites dans ce travail ne semble pas conduire à une sous estimation de la différenciation puisque le  $F_{ST}$  estimé à partir de locus RFLP, moins polymorphes entre les populations de Toulouse et Le Moulon après 10 générations d'évolution, est du même ordre de grandeur ( $F_{ST} = 0,03$ ) que les estimations obtenues dans notre expérimentation en G7 ( $F_{ST} = 0,02$ ) ou en G12 ( $F_{ST} = 0,04$ ) pour les mêmes populations.

#### → Biais pouvant affecter l'estimation du $Q_{ST}$ :

O'Hara et Merilä (2005) ont montré que l'estimation du  $Q_{ST}$  à partir d'un jeu de données incluant un faible nombre de populations ( $N < 20$ ) est peu précis en raison de la mauvaise estimation de la variance inter populations dans un tel contexte. Parmi tous les facteurs évoqués jusqu'à présent pouvant entraîner un biais dans l'interprétation de la différenciation observée aux caractères quantitatifs, il semble que ce point soit une des plus grosse faiblesse de notre dispositif puisque seules 3 populations sont considérées pour l'estimation de la différenciation.

Enfin, dans le cas d'un caractère non additif, les effets de dominance ou d'épistasie sont susceptibles d'affecter l'estimation du  $Q_{ST}$  (Whitlock 1999, Merilä et Crnokrak 2001, Lopez-Fanjul *et al.* 2003, Lopez-Fanjul *et al.* 2007). Ainsi Lopez-Fanjul *et al.* (2003) prédisent que la comparaison  $F_{ST}$  -  $Q_{ST}$  ne peut être utilisée pour détecter de la sélection que pour des caractères à base génétique purement additive puisque les effets de dominance et d'épistasie peuvent être à l'origine de la différence observée entre les deux mesures de différenciation. Les auteurs soulignent que cette prédiction est particulièrement problématique pour les caractères liés aux traits d'histoire de vie qui présentent généralement une part de variance non additive. Cependant quelques études théoriques (Le Corre et Kremer 2003, Goudet et Büchi 2006, Goudet et Martin 2007) et expérimentales récentes (Porcher *et al.* 2006) ainsi que la méta-analyse tendent à montrer que l'observation  $Q_{ST} > F_{ST}$  est une preuve robuste de sélection diversifiante (Leinonen 2008). Ainsi il semble clair que la précocité a été sélectionnée de façon divergente dans les populations de gestion dynamique. De même, malgré l'augmentation unidirectionnelle de hauteur comparée à la population parentale et à la population initiale, la hauteur a évolué de façon divergente dans les différents sites avec la

population de Vervins qui est significativement plus haute que les autres populations. En revanche, du fait des limitations de la comparaison  $Q_{ST} / F_{ST}$  exposées plus haut, nous ne pouvons évoquer la sélection homogénéisante pour expliquer l'absence de différenciation entre les populations pour le nombre et le poids de grains. Cependant ce résultat tend à montrer que toutes les populations évoluent vers le même niveau de productivité globale. Cette observation va dans le sens des résultats obtenus dans la thèse de J. David (1992, cf. Encadré IV-1).

#### **IV-4.2 Les forces de sélection agissant sur la précocité en relation avec les composantes de la valeur sélective**

L'analyse des gradients estimés à partir de la réponse à la sélection entre la génération 2 et la génération 12 de chaque population, tenant compte des covariances génétiques entre caractères, tend à montrer que la précocité ne serait pas sélectionnée de façon directe mais serait plutôt entraînée par la sélection agissant sur d'autres caractères. En raison du faible effectif évalué dans chacune des populations, le dispositif considéré est peu puissant et ne nous permet pas d'identifier de façon précise les caractères induisant l'évolution de la précocité observée dans chacune des populations considérées. Seule la hauteur montre des gradients de sélection importants dans toutes les populations traduisant le fait que ce caractère serait sélectionné de façon directe.

Les gradients de sélection estimés dans cette approche sont particulièrement faibles comparés à ceux qui sont classiquement observés dans les populations naturelles (Endler 1986 ; Kingsolver *et al.* 2001) ou ceux attendus en conditions contrôlées (Figure IV-3). Cependant l'approche utilisée ici donne une estimation de la force de sélection qui repose sur l'intensité de la réponse à la sélection entre 2 générations. Ainsi, les gradients estimés correspondent à la force de sélection agissant sur les caractères à chaque génération pour conduire à l'évolution observée au bout de 10 générations. Cette méthode suppose que les caractères ayant répondu à la sélection dans l'intervalle de temps considéré ont évolué de façon linéaire au cours des générations en subissant une force de sélection constante. Si cette hypothèse semble être valable pour la précocité de floraison, qui évolue de façon constante au cours des générations dans l'ensemble des populations pour l'intervalle de temps considéré (Figure IV-1), ceci n'est pas vrai pour la hauteur qui augmente très fortement entre la génération 2 et la génération 7 et reste stable entre la génération 7 et la génération 12. Les gradients de sélection estimés entre

la génération 2 et la génération 7 ont des valeurs bien supérieures en valeur absolue à ceux estimés entre la génération 2 et la génération 12 (de -0,202 pour le nombre de grains par épi dans la population de Vervins à 0,115 et 0,123 pour la hauteur dans les populations de Toulouse et Vervins). Cependant le sens de réponse à la sélection reste inchangé si l'on considère l'un ou l'autre des intervalles de temps donc les conclusions générales n'en semblent pas altérées.

L'estimation des gradients est obtenue à partir des variances et covariances observées entre caractères dans la population initiale considérée (ici génération 2). La validité des gradients dépend donc fortement de la qualité de détermination de l'état initial. De plus, l'extrapolation des gradients de sélection estimés sur une seule génération à la force de sélection s'exerçant sur plusieurs générations suppose que la matrice de variance covariance génétique est constante au cours des générations. La validité de cette hypothèse est discutable car la matrice peut évoluer sous le fait de la sélection elle-même ou peut dépendre de l'environnement dans lequel elle est mesurée.

L'ensemble de ces considérations nous oblige à considérer avec prudence les résultats obtenus à partir de l'analyse des gradients observés. Cependant l'observation selon laquelle la précocité ne serait pas sélectionnée de façon directe est concordante avec l'estimation des gradients attendus dans des conditions environnementales contrôlées. Dans les 2 expérimentations en conditions contrôlées pour la durée de la période de vernalisation, les individus les plus précoces sont sélectionnés de façon indirecte. Pour la condition « 4 semaines de vernalisation », la sélection de génotypes précoces s'explique par le fait que les génotypes dont les besoins en vernalisation ne sont pas satisfaits tendent à ne pas fleurir du tout ou à fleurir plus tard en pleine saison estivale et leur production de grains est très réduite. En revanche, pour la condition « 8 semaines de vernalisation », on peut considérer que les besoins en vernalisation de l'ensemble des plantes sont satisfaits. La sélection des plantes les plus précoces dans ce contexte, alors qu'elles ne sont pas sous sélection dans la condition au champ, s'explique par le fait que les plantes semées en février et sorties de vernalisation en avril ont produit beaucoup moins de végétation que les plantes semées en automne et qui ont passé l'hiver au champ. Ainsi dans les conditions 8S, les plantes n'ont pas le même stade physiologique que les plantes au champ et les plantes à cycle de vie court, capable d'initier rapidement la floraison sont favorisées.

Aucune sélection directe pour la hauteur n'est attendue dans les contextes environnementaux testés contrairement à ce qui est réellement observé dans les différentes populations. L'augmentation généralisée de la hauteur dans les différentes populations de gestion

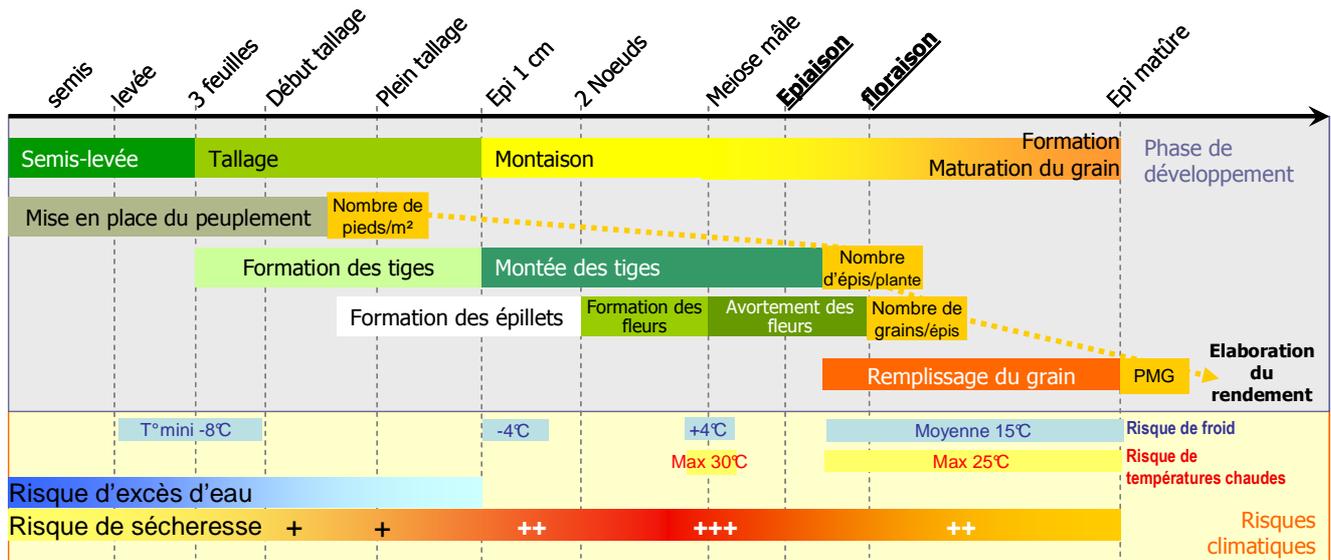
dynamique a souvent été mise en relation avec la bonne aptitude à la compétition des plantes hautes malgré leur productivité individuelle généralement plus faible. L'analyse des gradients de sélection attendus dans des conditions environnementales contrôlées a été effectuée en conditions de faibles densités et une expérimentation supplémentaire faisant varier les conditions de densité serait nécessaire pour mettre en évidence les conditions environnementales pour lesquelles une sélection pour la hauteur serait attendue.

L'analyse des gradients de sélection dans des conditions environnementales dans lesquelles on contrôle un paramètre (ici la durée de la période de vernalisation) permet de mettre en évidence des facteurs environnementaux ou agents de sélection pouvant expliquer la sélection observée. Ce type d'approche a notamment été utilisé pour estimer les gradients de sélection dans des conditions d'intensité lumineuse variables dans le but de mettre en évidence le rôle de la luminosité dans la sélection agissant sur des populations d'une plante herbacée (*Arrhenatherum elatius*) qui vit dans des milieux ouverts de type prairies et dans des milieux fermés en sous-bois (Petit et Thompson 1998). Plus récemment, Kingsolver *et al.* (2007) montrent une relation entre température et vitesse de croissance des larves de *Pieris rapae*, les températures plus fraîches tendant à sélectionner des larves à taux de croissance plus faibles. Cependant, il est clair que l'environnement conduisant à de la sélection dans les populations de blé se compose de divers facteurs tels que la température, la luminosité la pluviométrie, mais aussi les pressions pathogènes ou encore la densité de semis, et qui interagissent de façon complexe. Dans le cadre de l'étude de l'adaptation des populations, il serait intéressant d'effectuer des transplantations réciproques des populations dans les différents sites de culture pour mettre en évidence l'adaptation locale et tester si les individus, sur leur site d'origine, ont une meilleure valeur sélective que les individus des autres sites.

#### **IV-4.3. La précocité évolue en fonction des caractéristiques climatiques de chaque site**

L'analyse de la sélection dans des conditions environnementales contrôlées pour la durée de la période de vernalisation et les données de la littérature (Roux *et al.* 2006) montrent que les génotypes plus précoces sont sélectionnés lorsque les conditions de vernalisation sont incomplètes (conditions 4S) et dans une moindre mesure lorsque la durée de la période favorable à la reproduction est raccourcie (conditions 8S).

Figure IV-9 : Risques climatiques au cours des différentes phases d'élaboration du rendement chez le blé (Source : Robert *et al.* 1993).



Du fait de son importance agricole, les phases critiques de l'élaboration du rendement (production d'une parcelle estimée en terme de nombre ou poids de grains) sont bien identifiées chez le blé. Trois phases en liaison avec la date d'initiation florale ont été identifiées comme particulièrement sensibles aux perturbations climatiques (Figure IV-9) : la montaison, le stade de formation du pollen et la phase de remplissage du grain après épiaison. Les composantes du rendement étant des facteurs essentiels de la valeur sélective des populations, la sélection divergente de la précocité observée peut être mise en relation avec les caractéristiques climatiques des différents sites de culture et notamment la température (Figure IV-10). Dans le sud de la France, les plantes fleurissant de façon tardive sont susceptibles d'être exposées au stress hydrique et à la chaleur estivale au moment même de la phase cruciale de remplissage du grain. Ainsi les plantes à cycle de vie court arrivant à maturité de façon plus précoce tendent à être favorisées. Au contraire, dans le Nord de la France, les plantes initiant la floraison trop précocement peuvent être victimes de gelées tardives au Printemps, alors que se mettent en place les composantes de la valeur sélective à travers l'initiation des épillets conditionnant par la suite le nombre de grains. Ainsi dans le Nord, les génotypes tardifs à cycle de vie long, accumulant des ressources avant d'initier la reproduction lorsque les conditions du milieu sont meilleures, tendent à être favorisés.

Face aux contraintes environnementales contrastées, différentes stratégies de sensibilité aux facteurs environnementaux et de caractéristiques intrinsèques ont été adoptées dans les différentes populations considérées pour ajuster la floraison sur le cycle des saisons. Ainsi,

dans les populations les plus au Nord, pour lesquelles les variations inter saisons de la photopériode est la plus marquée (Le Moulon, Vervins, Figure IV-11), les génotypes insensibles à la longueur du jour disparaissent complètement tandis que les génotypes insensibles dans la population de Toulouse se caractérisent par une très grande précocité. Cependant même dans cette population, la photopériode est un signal important pour initier la floraison car, en génération 12, 40% des génotypes sont insensibles à la vernalisation mais  $\frac{3}{4}$  sont sensibles à la longueur du jour permettant aux plantes d'ajuster la floraison précocement une fois l'hiver passé. Dans la population du Moulon, les 2 signaux environnementaux semblent être des composantes majeurs déterminant la date de floraison puisque les génotypes sensibles à la fois à la longueur du jour et à la vernalisation sont très majoritaires en dernière génération.

Figure IV-10 : Evolution des températures au cours des mois de culture du blé (semis courant novembre, épiaison de mai à juin et récolte de juillet à août) dans les 3 sites considérés dans cet étude, obtenue à partir des températures moyennes mensuelles évaluées durant 30 années (1961 à 1990). Les courbes en pointillées indiquent les températures minimales et maximales de chaque mois durant cette période.

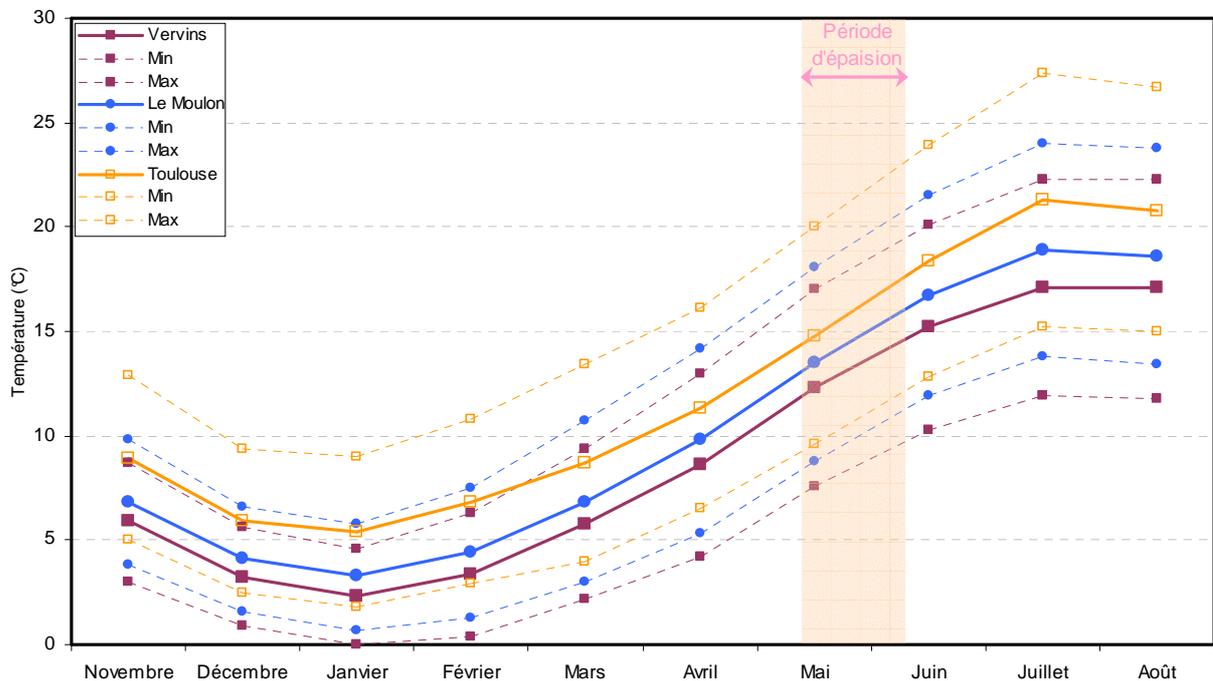
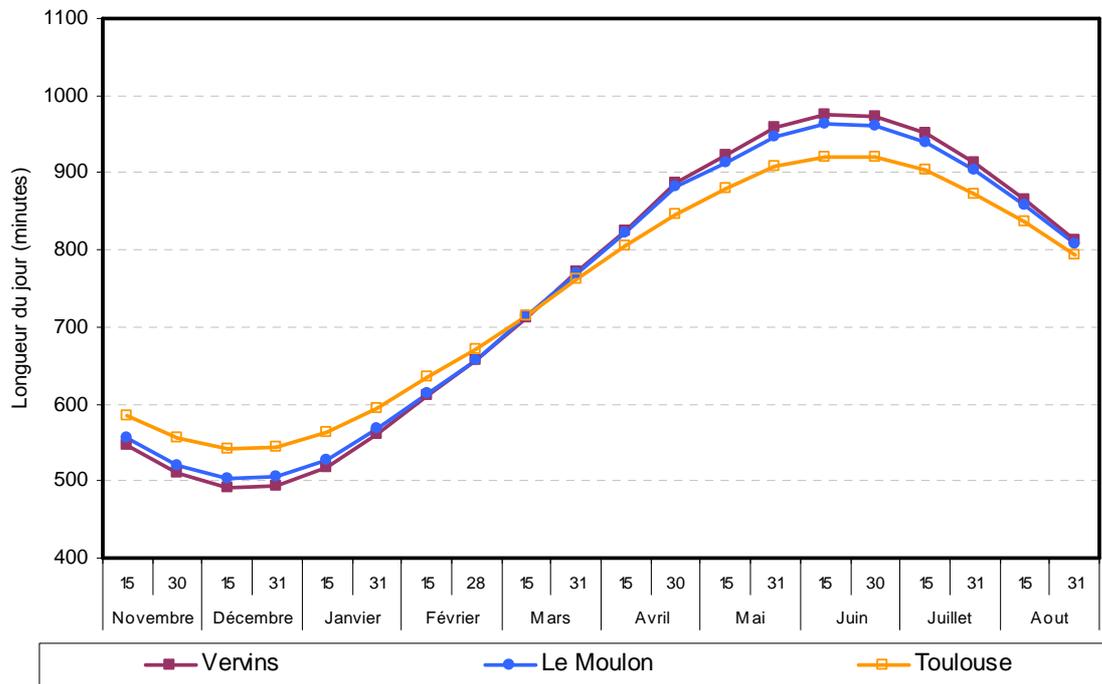


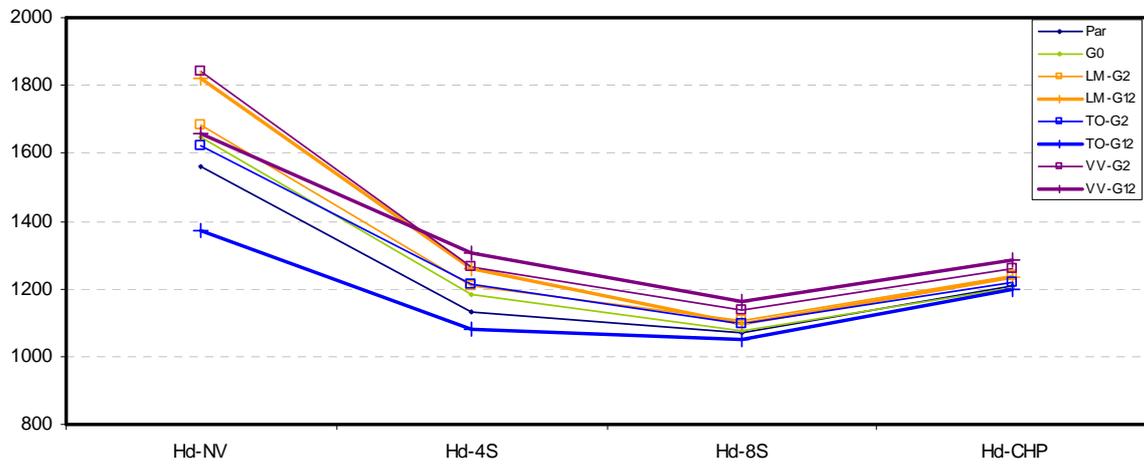
Figure IV-11 : Evolution de la longueur du jour au cours des mois de culture du blé (novembre à août) dans les 3 sites de culture des populations de gestion dynamique considérées dans cette étude.



#### Cas particulier de la population de Vervins :

L'étude des marqueurs microsatellites a montré que la population de Vervins a subi un fort goulot d'étranglement lors de son installation. Le très faible effectif efficace estimé entre la population initiale et la génération 2 a été mis en relation avec le fait que cette population est composée pour moitié de génotypes génétiquement très proches. On montre ici que ces génotypes sont caractérisés par une grande tardivité comparés à la population initiale et qu'ils sont sensibles à la vernalisation (Figures IV-1 et IV-7), suggérant que les génotypes tardifs sensibles à la vernalisation ont été très fortement sélectionnés dès les premières générations dans cette population (Figure IV-12). Cependant en dernière génération étudiée, une grande part des génotypes observés est de type Printemps, insensible à la vernalisation, d'après l'observation en conditions NV (absence de vernalisation) mais extrêmement tardifs dans les autres conditions observées suggérant que la précocité est essentiellement gouvernée par l'état intrinsèque des plantes. Dans les conditions environnementales de Vervins, on peut penser que les plantes sensibles au froid et à la photopériode auront leurs besoins satisfaits en début de printemps alors que des gelées tardives peuvent encore avoir lieu. Dans ce cas, de tels aléas climatiques tendraient à favoriser les génotypes dont le rythme de développement est intrinsèquement tardif et non dépendant des conditions climatiques.

Figure IV-12 : Valeur moyenne des épiaisons (en °J) dans les différentes populations en fonction des conditions de vernalisation. NV : absence de vernalisation, 4S : 4 semaines de vernalisation, 8S : 8 semaines de vernalisation et CHP : épiaison au champ.



## IV-5 Conclusions

Malgré les grandes tendances qui se dégagent dans les différentes populations, en ce qui concerne les stratégies de sensibilité aux facteurs environnementaux et les caractéristiques intrinsèques, pour initier la floraison lorsque les conditions du milieu sont favorables, on constate que les populations présentent également une grande diversité à l'intérieur même des populations au bout de 12 générations d'évolution, en particulier pour les populations les plus extrêmes. L'évolution linéaire de la précocité au cours des générations dans les populations de Toulouse et Vervins semble montrer que les populations n'ont pas encore atteint leur optimum de précocité au bout de 12 générations. Ce qui n'est par exemple pas le cas pour la hauteur pour laquelle il a été montré que la relaxe de sélection observée pour la hauteur entre la génération 7 et la génération 12 est principalement liée à la perte de diversité pour les gènes majeurs de nanisme dans les différentes populations (L'allèle nain du gène majeur *Rht1* est perdu dans la population du Moulon après 17 générations d'évolution ; Raquin *et al.* 2008). Pour la précocité floraison gouvernée par un grand nombre de gènes en interaction, le suivi de l'évolution de la diversité des gènes impliqués dans le caractère devrait nous permettre de mieux comprendre comment le maintien de cette diversité est encore possible au bout de 12 générations d'évolution et de sélection pour la précocité (Chapitre V).

Enfin, il est intéressant de noter que dans les populations les plus extrêmes, les dates de floraison les plus extrêmes ont lieu pour les génotypes insensibles aux températures froides (*i.e.* ne nécessitant pas de températures froides pour fleurir) et dont la floraison est majoritairement gouvernée par l'état intrinsèque de la plante et la longueur du jour. Ainsi, dans la population de Toulouse, les génotypes les plus précoces sont à la fois insensibles à la photopériode et à la vernalisation, tandis que dans la population de Vervins, les génotypes insensibles et présentant une tardivité intrinsèque sont ceux qui fleurissent le plus tard. L'insensibilité aux températures froides permet de ne pas dépendre de certains états environnementaux (aléas climatiques, gelées tardives, sécheresse estivale précoce) qui peuvent être variables d'une année sur l'autre. On peut penser que du fait des conditions climatiques de plus en plus variables, les génotypes les mieux adaptés dans les différents environnements seront ceux qui présentent une précocité intrinsèque adaptée ou dont la floraison est gouvernée par des signaux environnementaux stables (longueur du jour).

## *Chapitre 5*

# *EVOLUTION CONJOINTE DES CARACTERES QUANTITATIFS, DES MARQUEURS MICROSATELLITES ET DES GENES CANDIDATS DE PRECOCITE*

# Chapitre V: Evolution conjointe des caractères quantitatifs, des marqueurs microsatellites et des gènes candidats de précocité

## V-1 Introduction

L'objectif de ce chapitre est de suivre l'évolution de la diversité de gènes potentiellement impliqués dans la précocité sachant que ce caractère est sélectionné de façon divergente dans les populations. Pour cela, je cherche à identifier si les variations des gènes étudiés sont associées avec la variation pour la précocité de floraison dans les populations. Puis je cherche à identifier si les gènes impliqués présentent des signatures de sélection. Dans un premier temps, j'étudie l'évolution des 3 copies du gène majeur, *VRN-1*, identifié comme impliqué dans la voie de la vernalisation, en relation avec le caractère de précocité estimé en l'absence de vernalisation (HdNV). La distribution bimodale de ce caractère permet de distinguer les génotypes de type « Hiver » nécessitant une période de froid pour fleurir et donc sensibles à la vernalisation, des génotypes de type « Printemps » insensibles. La mise en relation de ce caractère, pouvant être traité comme un caractère quantitatif, avec l'évolution d'un gène majeur, mais néanmoins présent en 3 copies, permet d'étudier de façon plus approfondie la relation entre le phénotype et le génotype dans un cas relativement simple. Dans un deuxième temps, l'évolution du polymorphisme nucléotidique de 6 gènes de floraison est mise en relation avec l'évolution de la précocité de floraison estimée au champ (HdCHP). Cette mesure de précocité est une variable plus intégrative que celle mesurée précédemment puisqu'elle représente la condition « normale » de culture du blé sous nos latitudes (semis d'automne et épiaison en avril), et donc des populations de gestion dynamique (à la densité près), les plantes épiaient une fois que leurs besoins en photopériode et en vernalisation sont satisfaits. Dans cette étude, l'accent est mis sur les interactions observées entre gènes dans la construction du caractère et l'importance de la relation entre les gènes dans la réponse à la sélection. Ce chapitre se présente sous la forme de deux articles, le premier est publié dans la revue « Molecular Ecology » (Rhoné *et al.* 2008) et le second est en préparation pour la revue « Evolution ».

## **V-2 Relations entre l'évolution du gène majeur de la voie de la vernalisation et l'évolution de la sensibilité à la vernalisation**

Dans cet article, l'évolution de la diversité du gène *VRN-1* impliqué dans la réponse à la vernalisation chez le blé a été mise en relation avec l'évolution de la précocité estimée en condition contrôlée, en l'absence de vernalisation. Dans la littérature, le gène *VRN-1* a été montré comme impliqué dans l'architecture du caractère de précocité chez le blé et en particulier, dans la voie de la vernalisation et la détermination du type « Hiver » (H) ou « Printemps » (P). Du fait la nature hexaploïde du blé ( $2n = 42$ , génomes AABBDD), *VRN-1* est présent sous forme de 3 copies fonctionnelles chez le blé (*VRN-1A*, *VRN-1B* et *VRN-1D*). Ces 3 copies présentent des relations d'épistasie entre elles : la présence d'un allèle dominant de type P est suffisant pour conférer un caractère « P » au phénotype le rendant de ce fait insensible à la vernalisation. Deux régions polymorphes sur la copie A du gène, une sur la copie B et une sur la copie D, ont été génotypées pour les 3 populations expérimentales de blé aux 3 générations étudiées. Les variations des 3 copies sont très fortement associées aux variations pour le caractère H/P des individus étudiés et expliquent près de 80% de la variation pour la précocité en l'absence de vernalisation. Quelle que soit la population considérée, de fortes variations génotypiques associées à des changements phénotypiques rapides sont observées dès les 7 premières générations d'évolution. Dans chaque environnement, des combinaisons alléliques différentes sont favorisées permettant le maintien d'un fort niveau de diversité sur l'ensemble des populations. De façon plus surprenante, dans 2 des 3 populations étudiées, l'apparition de nouveaux allèles par mutation ou par migration, probablement couplée à une sélection variable dans le temps ou à une sélection fréquence-dépendante, favorisant les génotypes rares, permet le maintien d'une grande diversité intra population.

En se basant sur des populations expérimentales suivies sur plusieurs générations, cette étude offre de nouvelles avancées dans la compréhension des bases génétiques de l'adaptation. Premièrement, l'évolution rapide des fréquences alléliques du gène majeur de vernalisation entre les premières générations observées va dans le sens du modèle de Fisher/Orr qui prédit des variations phénotypiques importantes aux premières générations impliquant des gènes à effet fort, suivi d'un ajustement plus fin aux générations suivantes. Par ailleurs, les

interactions épistatiques observées entre copies reposent sur des interactions entre régions régulatrices des différentes copies du gène : dans les blés de types H, l'expression de *VRN-1* est contrôlée par des répresseurs agissant sur des régions régulatrices du promoteur et du premier intron. Pour le blé, le caractère Hiver est un caractère ancestral. Une mutation altérant l'un des sites de reconnaissance des répresseurs du gène dans l'une de ces copies conduit à l'expression de *VRN-1* dans les blés non vernalisés et donc à un phénotype de type P. Du fait de ces interactions entre copies, la sélection pour le caractère H/P tend à opérer sur des combinaisons alléliques particulières créant des déséquilibres de liaison entre les copies du gène situées sur des chromosomes différents. Un même phénotype pouvant être obtenu par différentes combinaisons, la sélection peut favoriser divers génotypes multilocus dans un même contexte environnemental permettant ainsi le maintien de la diversité dans les populations.

# Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations

BÉNÉDICTE RHONÉ, CARINE REMOUÉ, NATHALIE GALIC, ISABELLE GOLDRINGER and ISABELLE BONNIN

UMR de Génétique Végétale, INRA/CNRS/AgroParisTech/Univ. Paris XI, Ferme du Moulon, 91190 Gif sur Yvette, France

## Abstract

Experimental populations evolving under natural selection represent an interesting tool to study genetic bases of adaptation. Evolution of genes possibly involved in adaptive response can be followed together with the corresponding phenotypic traits. Using experimental populations of hexaploid wheat, we studied the evolution of flowering time, a major adaptive trait that synchronizes the initiation of reproduction and the occurrence of favourable environmental conditions. During 12 generations, three populations were grown in contrasted environments (Vervins North France, Le Moulon near Paris, Toulouse South France) under the influence of natural selection, drift, mutation and recombination. Evolution of diversity at the major gene *VRN-1* involved in wheat vernalization response has been analysed jointly with earliness estimated in controlled conditions. Whatever the population, rapid phenotypic changes as well as parallel genotypic variations were observed in the first seven generations, probably as the result of selection acting on this major gene which explains 80% of the trait variation overall. Different allelic combinations at physically unlinked copies of *VRN-1* located on distinct genomes (A, B and D) were selected between populations. As theoretically expected, due to population differentiation, a high level of genetic diversity was maintained overall in generation 12. Surprisingly, in two populations out of three, the emergence of new alleles by mutation or migration, coupled with temporal variable selection or frequency-dependent selection, allowed to maintain within-population diversity despite local genetic drift and natural selection. This result may plead for an evolutionary approach of wheat genetic resource conservation.

*Keywords:* adaptation, differentiation, experimental populations, flowering time, growth habit, *Triticum aestivum* L.

Received 27 May 2007; revision received 28 September 2007; accepted 26 October 2007

## Introduction

Understanding mechanisms involved in local adaptation linked to spatial or temporal heterogeneity of selection has become an important issue of evolutionary and conservation biology. Recent progress in functional genomics and molecular genetics allowed the genetic dissection of adaptive traits and clarified the molecular architecture of adaptation (Schlötterer 2002). However, how selection on phenotypes shapes molecular genetic variation (number and effects of genes, type of mutations involved in

selection response) is still unclear. Many recent reviews encouraged experimental studies aiming at better understanding the genetic basis of adaptation by combining the study of the evolution of adaptive traits jointly with their underlying molecular bases (Orr 2005; Ehrenreich & Purugganan 2006).

Experimental populations of model species evolving under natural selection represent an interesting tool to study mechanisms of adaptation. Since several parameters (such as initial composition of the population and population size) can be controlled, useful genetic parameters can be precisely estimated and the evolution of genes possibly involved in adaptive response can be followed together with the corresponding phenotypic traits. In experimental wheat populations grown in different climatic

Correspondence: B. Rhoné, Fax: +33 (0)1 69 33 23 40; E-mail: rhone@moulon.inra.fr

conditions, Goldringer *et al.* (2006) found a north–south differentiation for flowering time by comparing populations from different sites after 10 generations of growth, which was interpreted as a response to local climatic conditions. Flowering time is a major adaptive trait in annual plant species because it synchronizes the initiation of reproduction with the seasons ensuring reproductive success under favourable environmental conditions (Remington & Purugganan 2003; Roux *et al.* 2006). Thus, studying the genetic bases (number and diversity of allelic combinations) underlying the phenotypic differentiation found between the experimental wheat populations appears as an efficient approach to better understand the genetic bases of adaptive response.

Flowering time is a complex trait controlled both by the internal states of the plant and environmental conditions, particularly temperature and light. Molecular regulation of the floral transition has been widely studied in *Arabidopsis thaliana*. The study of this model species highlights the complexity of the regulation network of this trait involving more than 80 genes acting on four genetic pathways (for a review, see Mouradov *et al.* 2002; Putterill *et al.* 2004). However, much less is known about flowering control of other plants species. The mechanisms of flowering time control seem to be well conserved among plants in particular for photoperiodic pathways that involved orthologous genes with similar function in monocots and dicots (Andersen *et al.* 2004; Yan *et al.* 2006).

For most temperate crops, winter frost is a major environmental signal that regulates flowering time (Loskutov 2001; Goldringer *et al.* 2006). In cereals such as wheat, barley and oat, both winter genotypes (which need a period of exposure to low temperature called vernalization to initiate flowering) and spring genotypes (insensitive to vernalization) exist, indicating a large range of adaptive responses to vernalization (Roux *et al.* 2006). Wheat is also classified as a long day plant because most of the genotypes flower when exposed under long photoperiod. Under temperate climates, this crop is usually sown in autumn and the requirement of a long exposure to cold and a long photoperiod ensure the plant initiates flowering under favourable conditions in the spring season. Numerous analyses of wheat segregating populations, grown in different conditions of day length and temperature have revealed three important pathways for flowering time: vernalization (involving *VRN* genes), photoperiod (*PPD* genes) and earliness per se (*EPS* genes; Law & Worland 1997). Due to the allohexaploid nature of the cultivated wheat genome ( $2n = 42$ , genomes AABBDD), each gene usually possess three homeologous copies. Recombination between the three genomes is prevented by the activity of pairing homeologous loci (Dvorak & Lukaszewski 2000). For example, the *VRN-1* gene is present in three copies within the genome, *VRN-1A*, *VRN-1B* and *VRN-1D*. These

three copies show epistatic interactions: a dominant spring allele at any one of these loci is sufficient to confer a spring growth habit (Stelmakh 1987). Only recently, the first molecular information was made available for flowering genes in wheat, when Yan *et al.* (2003) succeeded in isolating the *VRN-1* gene by positional cloning in the einkorn wheat (*Triticum monococcum*,  $2n = 14$ , genome  $A^m A^n$ ), a wild diploid species related to cultivated wheat. This gene was found to be similar to an *Arabidopsis* meristem identity gene, *APETALA1* (*API*). In hexaploid wheat, *VRN-1* is expressed in leaves and apices of spring wheat and in vernalized winter wheat (Yan *et al.* 2003). Deletions in the promoter region of *VRN-1A* appeared to be linked to dominant spring growth habit (Yan *et al.* 2004), as well as deletions of 4–6 kb length in the first large intron (9 kb length) of *VRN-1B* and *VRN-1D* suspected to contain regulatory site. Moreover, Sherman *et al.* (2004) provided a polymerase chain reaction (PCR) marker for growth habit based on a substitution in the seventh exon of *VRN-1A*, which can be used to determine the allelic state at *VRN-1A* in germplasm collections. Until now, these four polymorphic regions have not been simultaneously characterized in the same genotypes in relation to their growth habit. However, regulatory elements such as those found in the promoter region and in the first intron may play important roles in adaptation by controlling the temporal and spatial expression of genes (Purugganan 2000; Wray 2007). Thus, the three copies of the *VRN-1* gene constitute a model system for a parallel study of genotypic and phenotypic adaptive response. In the present work, we study the evolution of growth habit frequencies jointly with the evolution of polymorphisms at the major gene *VRN-1* governing this earliness component in three experimental populations of wheat sampled at three generations (2, 7 and 12). We first clarify the genetic architecture of growth habit and the relationship existing between the three homeologous copies of the *VRN-1* gene. We then identify whether specific allelic combinations were selected in the different environments, and we compare the evolution of genotypic and phenotypic diversity in the different populations.

## Materials and methods

### *Plant material*

The experimental populations studied here belong to a dynamic management programme of wheat genetic resources. Dynamic management (DM) is an evolutionary approach for genetic resource conservation aiming at maintaining genetic diversity of crop species by growing successive generations of an initial composite population in several contrasted environments, without intentional human selection. These experimental populations are submitted to the evolutionary pressures encountered by

natural populations such as natural selection and genetic drift, and they are used as a model for studying the mechanisms involved in local adaptation to environmental conditions. While each population, in the absence of migration or mutation, is expected to lose part of its initial variability by genetic drift and selection, diversity is expected to be maintained overall. The DM population of wheat was created by four successive generations of crosses among 16 diverse homozygous genotypes with contrasted values for earliness, height, resistance levels and other fitness-related traits. The four consecutive crosses aimed at maximizing recombination between the different parental alleles before starting the selection experiment. After three generations of multiplication to increase seed amount, the composite population obtained called G0 constituted the initial population of the DM programme. In 1984, seed samples of G0 were sown in seven sites of a French experimental network. Since then, each population has been cultivated every year without human selection and without migration between populations. Each year, the whole plot was harvested in bulk, a random sample of at least 5000 seeds was sown in a 100 m<sup>2</sup> plot isolated from other wheat cultures to constitute the next generation and 500–1000 g of seeds were stored in a cold room (for more details see Goldringer *et al.* 2006 and Rhoné *et al.* 2007).

Here, we studied the spatial and temporal evolution of growth habit and *VRN-1* genotypes in the DM populations using: (i) the 16 parental lines; (ii) 36 inbred lines of the initial population G0; and (iii) 540 genotypes (60 genotypes for each population at each generation) taken from the generation 2 (G2 grown in 1986), generation 7 (G7; 1991), and generation 12 (G12; 1996) of the populations that have been cultivated under extensive conditions (low nitrogen input and no fungicide) in Vervins (VV; 49.5°N, 3.5°E), in Le Moulon (LM; 48.4°N, 1.4°W) and in Toulouse (TO; 43.4°N, 1.3°E). The G0 seeds used in this study were inbred lines derived by at least three generations of selfing from the initial population. For the 540 genotypes of the DM populations, seeds were sampled from the seed stock stored in a cold room at each generation. Each plant grown from those seeds was identified as a single genotype. The 592 genotypes of this experiment were sown at Le Moulon in 2004 at low density in 6 rows × 1.35 m plots. We took a leaf sample for DNA extraction and we used the selfed progenies of each genotype to determine their growth habit. Since wheat is mainly selfing (estimated outcrossing rate of 4–6% in this population, Enjalbert & David 2000), selfed progenies of each genotypes were highly homogenous.

#### Experimental design

In April 2005, six seeds per genotype were evenly sown over a 0.70 m distance in the nursery at Le Moulon. All

genotypes were randomized in a 72 × 1.40 m plot with two genotypes per 1.40 m row. The seedlings were grown during spring and summer seasons so that they were not exposed to vernalizing temperature (< 10 °C) and were grown under natural long photoperiod. For each plant, the date of heading was recorded when half of the main ear had emerged from the flag leaf. The heading date was transformed into sums of degree-days (sums of the mean temperature per day) from germination to heading. The experiment was ended on 1 August (1752 °C day cumulated). A value of 2000 °C day was given to plants that did not reach heading at this deadline. The mean value of heading date over the six plants was calculated for each genotype and denoted HdNV (heading date without vernalization). Given that the HdNV frequency showed a bimodal distribution, the value of 1500 °C day (see Fig. S1a, Supplementary material) was chosen to separate winter growth habit genotypes from spring growth habit genotypes.

#### Molecular analyses

For each plant, total DNA was extracted from 200 mg of young leaves following a protocol derived from the DNeasy 96 Plant kit (QIAGEN).

Four *VRN-1* polymorphic sites were genotyped in this study: (i) duplication, insertion and deletion in the promoter of *VRN-1A* (denoted *VRN-1A<sub>pr</sub>* in the following) revealed by Yan *et al.* (2004), (ii) a substitution in the seventh exon of *VRN-1A* (*VRN-1A<sub>ex7</sub>*) revealed by Sherman *et al.* (2004), (iii) a 4-kb deletion in the first intron of *VRN-1B* (*VRN-1B<sub>int1</sub>*) and (iv) a 4-kb deletion in the first intron of *VRN-1D* (*VRN-1D<sub>int1</sub>*) revealed by Fu *et al.* (2005). For all the *VRN-1* polymorphic sites, PCR conditions and PCR product digestion protocols were the same as defined by the authors. To detect variations at *VRN-1A<sub>pr</sub>*, forward primers were modified with an M13 extension according to Boutin-Ganache *et al.* (2001). PCR amplifications were performed in the presence of fluorescent-labelled M13 tailed. The amplification products, loaded on 6.5% denaturing polyacrylamide gels, were analysed on a LI-COR automated DNA sequencer (LI-COR Biosciences). The amplification fragments sizes were assessed using the ONEDSCAN software (version 2.03, Scanalytics). The variations at *VRN-1A<sub>ex7</sub>* (CAPS marker) and at *VRN-1B<sub>int1</sub>* and *VRN-1D<sub>int1</sub>* (presence or absence of deletions), were revealed by migration on 2% and 0.8% agarose gels, respectively, and visualized with UV light.

Using migration on acrylamid gel to study *VRN-1A<sub>pr</sub>* variation, instead of less resolutive agarose gel as in Yan *et al.* (2004) allowed us to reveal two new alleles (denoted *vrn1A<sub>pr</sub>-3* and *vrn1A<sub>pr</sub>-4*) not described previously in the literature. Those new alleles were sequenced. The amplified fragments were purified using the QIAquick PCR

purification kit (QIAGEN). The  $vm1A_{pr-4}$  allele was directly sequenced by GENOME express in both forward and reverse directions from the purified PCR product of four genotypes. The allele  $vm1A_{pr-3}$  showed two bands after migration on polyacrylamide gel. The two fragments were ligated into pGEM-T Easy vector (Promega) and cloned into *Escherichia coli* DH5 $\alpha$  chemically competent cells (Invitrogen). To identify the clones carrying the high fragment and clones carrying the low fragment of the allele, PCR amplification was performed on white colonies using labelled M13 primers and the PCR products were separated on a LI-COR automated DNA sequencer. The sequences were produced from the PCR product of each fragment obtained from clones of four genotypes. One clone per band and per genotype was sequenced in both forward and reverse directions. All sequences were analysed using the STADEN software package version 1.6 (Staden *et al.* 1998).

#### Data analyses

*Genotypic data analyses.* Allelic frequencies were estimated for each population and each generation. Genetic diversity was calculated from the allelic frequencies using the unbiased estimator  $H_e$  developed by Nei (1987). The linkage disequilibrium (LD, nonrandom associations of alleles at two different loci) was estimated between loci using the common correlation coefficient ( $R$ ) and between alleles at different loci using the correlation coefficient ( $R_{ij}$ ). They were calculated with LINKDIS program available in the GENETIX package (Belkhir *et al.* 2004).

G0 was considered as the initial state for studying temporal evolution within the populations (Goldringer *et al.* 2006). Therefore, temporal differentiation was estimated for each site between G0 and the three generations G2, G7 and G12. Global temporal (in one site, across generations) and spatial (at one generation, among sites) differentiations were estimated for each locus separately and overall using Weir & Cockerham (1984)  $F$ -statistics estimator  $\theta$  calculated by the software program FSTAT version 2.9.3.2 (Goudet 2002). Significance of  $\theta$  values was tested using permutations of genotypes among samples to obtain a randomised data set. The log-likelihood statistic  $G$  was used to classify contingency tables (alleles  $\times$  samples) obtained from randomised data sets as implemented in FSTAT (Goudet *et al.* 1996). In this procedure, loci are weighted according to their 'information' content to combine individual locus  $P$  values. For instance,  $P$  values for very polymorphic loci have a higher weight than those for nearly monomorphic loci. Finally, adjusted  $P$  values for  $\theta$  were obtained applying Bonferroni correction for multiple comparisons. Pairwise tests of temporal and spatial differentiation between populations were also carried out by this process.

Given that strong linkage disequilibria among loci are expected in selfing species, multilocus data were analysed. Using the Bayesian ELB algorithm from ARLEQUIN 3.01 software (Excoffier *et al.* 2005), the most probable allelic associations at the four loci of *VRN-1* were reconstructed. As heterozygous are rare in wheat, we were able to identify the most probable association of alleles (haplotypic phase) in the two gametes constituting a genotype. Haplotypic data were then treated as a unique locus with a number of alleles equal to the number of haplotypes. Haplotype frequencies, Nei diversity  $H_e$  and spatio-temporal differentiation were estimated as described above. In order to compare spatial and temporal evolution of *VRN-1* haplotypes simultaneously, hierarchical analysis of molecular variance (AMOVA) was performed in ARLEQUIN 3.01 to partition total genetic variation among populations (Le Moulon, Toulouse and Vervins) and among generations within population (G2, G7, G12), using haplotype frequencies.

*Phenotypic differentiation analyses.* Spatial and temporal genetic differentiation for earliness was estimated using both quantitative (heading date HdNV) and qualitative (growth habit type) data.

Considering spatial and temporal evolutions simultaneously, analyses of variance were performed on HdNV using the GLM procedure of the SAS software package (SAS Institute 2000) with the following model:

$$Y_{ijk} = \mu + pop_i + gene(pop)_{ij} + \varepsilon_{ij} \quad (\text{eqn 1})$$

where  $pop_i$  is the effect of population  $i$  (Moulon, Toulouse or Vervins), and  $gene(pop)_{ij}$  is the effect of generation  $j$  (G2, G7 or G12) nested in the population  $i$ . Both effects were fixed. For each significant effect, adjusted means for the other effects in the model (*LS-Means*) were estimated. Normal distribution of the residual terms in the ANOVA was tested using the UNIVARIATE procedure of the SAS software.

Growth habit frequencies were estimated for each population at each generation, and exact tests of differentiation were performed between populations (or generations) and overall, using the GENEPOP software version 3.4 (Raymond & Rousset 1995). The level of spatio-temporal differentiation for winter/spring growth habit type was quantified using Weir and Cockerham  $\theta$  estimator in FSTAT version 2.9.3.2 (Goudet 2002).

*Relationship between VRN-1 polymorphism and earliness variations.* Given that *VRN-1A<sub>pr</sub>* was located 12 kb upstream *VRN-1A<sub>ex7</sub>* and was always in strong linkage disequilibrium with this locus, two-locus haplotypes were reconstituted and thus used to test the effect of the *VRN-1A* copy. We estimated the genetic contribution ( $R^2$ ) of the *VRN-1* polymorphisms on the HdNV variation considering the following ANOVA model:

$$Y_{ijk} = \mu + VRN-1A_i + VRN-1B_j + VRN-1D_k + (VRN-1A \times VRN-1B)_{ij} + \epsilon_{ijk} \quad (\text{eqn 2})$$

with  $VRN-1A_i$ ,  $VRN-1B_j$ ,  $VRN-1D_k$  the effect of the three  $VRN-1$  copies A, B and D. The interaction between  $VRN-1A$  and  $VRN-1D$  and between  $VRN-1B$  and  $VRN-1D$  were not tested since the genotypic classes were too unbalanced. For each significant effect, means of HdNV (*HdNV-Means*) were estimated.

In order to test for the genetic contribution of each genome on HdNV corrected for the contribution of the other genomes, partial  $R^2$  were estimated from model 2' (same as model 2 without the interaction term  $VRN-1A \times VRN-1B$ ) as follows:  $R^2_{part-X} = SS_X / SS_T$  where  $SS_X$  is the sum of squares calculated for the effect of the  $VRN-1X$  copy and  $SS_T$  is the total sum of square.

## Results

### *Allelic variation at the VRN-1 copies*

The three alleles of  $VRN-1A_{pr}$  described in Yan *et al.* (2004) were present within the 16 parents and the DM populations: the  $Vrn1A_{pr}-1$  allele, associated with spring growth habit and presenting a duplication of the promoter region; the  $vrn1A_{pr}-2$  allele, ancestral form of the gene and associated with winter growth habit and the  $Vrn1A_{pr}-5$  allele associated with spring growth habit and presenting a 20-bp deletion in a TC-repetitive region. However, two additional alleles (called  $vrn1A_{pr}-3$  and  $vrn1A_{pr}-4$ ) not described previously in the literature and absent in the parents were found in the population of Vervins since generation 2.

The allele  $vrn1A_{pr}-3$  showed two bands after migration on polyacrylamide gel. As checked by using a  $vrn1A_{pr}-2$  genotype as a control, one of the two fragments migrated at the same position as the allele  $vrn1A_{pr}-2$  ('high fragment') and the other migrated at a lower position ('low fragment'). Given that the lower band was never encountered alone, and that duplication of the promoter region was already observed (allele  $Vrn1A_{pr}-1$ , Yan *et al.* 2004), this result was interpreted as a duplication of this region. Moreover, considering that the two bands of  $vrn1A_{pr}-3$  are different alleles in heterozygous plants and not a duplication would involve a high level of heterozygote excess overall (frequency of 0.12 in G12 in the population of Vervins), which is unlikely in a selfing species such as bread wheat. After estimation of the size of the two fragments on the polyacrylamid gel, we expected a difference of two base pairs between the two sequences. Cloning of the PCR products of four genotypes from Vervins allowed us to isolate the two fragments in separate clones. The sequencing of the two bands revealed that both fragments had the same sequence as the allele  $vrn1A_{pr}-2$ , except for one SNP in the lower band (see Fig. S2, Supplementary material,

GenBank Accession nos EU127902 and EU127903). To explain the difference of migration between the two fragments despite the absence of length polymorphism, we tested whether it might be due to a highly stable conformation within one of the two fragments maintained despite denaturing conditions of migration (denaturing 5 min at 95 °C in formamide, migration at 45 °C on gel containing urea). For that, we increased the temperature of migration to 60 °C. In such extreme conditions of migration, the two fragments migrated at the same position. Moreover, we used the MFOLD software version 3.2 (Zucker 2003) to visualize the most probable conformation of DNA fragment estimated from the sequence in the two conditions of migration (45 °C and 60 °C). At 45 °C, the 'low fragment' with a G showed a hair pin in its conformation that was absent in the 'high fragment' with a C. This hair pin was observed in a C- and G-rich region (Fig. S2). At 60 °C, the conformation of the two fragments changed compared to the previous condition; the hair pin observed for the 'low fragment' is reduced and conformations of the two fragments are similar. These two observations suggested that the difference of migration observed between the two bands presenting no differences in length could be due to a difference of conformation during migration.

The  $vrn1A_{pr}-4$  allele fragment sequenced in four genotypes from Vervins (Fig. S2, GenBank Accession no. EU127901) showed the same 20-bp deletion in the TC-repetitive region of the 5' UTR as the  $Vrn1A_{pr}-5$  allele. The length polymorphism observed between these two alleles after migration on acrylamide gel was due to: (i) an insertion in a polyC near the zone identified as a host direct duplication (HDD) region by Yan *et al.* (2004), and (ii) a 2-bp insertion within this HDD region (Fig. S2).

The two alleles of  $VRN-1A_{ex7}$  and of  $VRN-1B_{int1}$  described in Sherman *et al.* (2004) and in Fu *et al.* 2005), respectively, were present within the 16 parents and the DM populations.  $VRN-1D_{int1}$  was not polymorphic in the parental population; the 16 parental lines carried the allele described as associated with a winter growth habit. However, the spring allele with a large deletion (Fu *et al.* 2005) appeared in the initial DM population G0 and was found at very low frequencies (< 0.05) in all the studied generations of the Toulouse population. Sequence analysis of this region was not performed due to the large size of the region concerned by the mutation (9 kb intron).

### *Spatio-temporal evolution of VRN-1 gene*

**Temporal evolution.** In the Le Moulon population, the frequencies of all the studied polymorphic loci evolved slowly across generations. Winter alleles at  $VRN-1A_{pr}$  and at  $VRN-1A_{ex7}$  remained at a high frequency (both at 0.833 in G12), while the  $vrn1B_{int1}-1$  winter allele increased progressively until fixation. This population was found

to be monomorphic for *VRN-1D<sub>int1</sub>*. No significant temporal evolution was found in this population ( $\theta = -0.003$ ;  $P = 0.06$ ), whereas the genetic diversity tended to decrease progressively across generations (from 0.20 in G2 to 0.11 in G12).

For the Vervins population, two new alleles *vrn1A<sub>pr</sub>-3* and *vrn1A<sub>pr</sub>-4* of the promoter appeared in G2 and increased in frequency until G12, replacing *vrn1A<sub>pr</sub>-2* and *Vrn1A<sub>pr</sub>-5* alleles. Both winter alleles from *VRN-1A<sub>ex7</sub>* and from *VRN-1B<sub>int1</sub>* strongly decreased across generations (from 0.97 in G2 to 0.51 in G12 for *vrn1B<sub>int1</sub>-1*,  $P < 0.0001$ ), whereas *VRN-1D<sub>int1</sub>* was not polymorphic in this population. The rapid evolution of allelic frequencies and the appearance of new alleles in the population led to an increase of genetic diversity between G2 and G7 (from 0.21 to 0.40). Diversity remained then stable between G7 and G12. Significant temporal differentiation was found in Vervins overall ( $\theta = 0.191$ ;  $P = 0.0002$ ). Variations at both *VRN-1A* and *VRN-1B* contributed to the strong evolution of the population (between G0 and G12:  $\theta = 0.397$ ,  $P = 0.005$  for *VRN-1A<sub>pr</sub>*;  $\theta = 0.312$ ,  $P = 0.005$  for *VRN-1A<sub>ex7</sub>* and  $\theta = 0.425$ ,  $P = 0.005$  for *VRN-1B<sub>int1</sub>*).

In the Toulouse population, allelic frequencies of the two polymorphic regions of *VRN-1A* evolved in an erratic way. At the first generation, winter allele frequencies decreased significantly for the two regions until G7 ( $P < 0.0001$  between G0 and G7), whereas these winter alleles increased in frequency between G7 and G12 ( $P < 0.05$ ). The frequency of the spring allele *Vrn1B<sub>int1</sub>-2* increased from 0.03 in G2 to 0.35 in G12, whereas the spring allele *Vrn1D<sub>int1</sub>-2*, absent in the parents but present in G0, was maintained at a low frequency from 0.01 in G2 to 0.03 in G12. As for the Vervins population, gene diversity increased rapidly between generations 2 and 7 (from 0.24 to 0.38) and then remained stable until generation 12. Significant temporal differentiation was found in the Toulouse population overall ( $\theta = 0.074$ ,  $P = 0.0002$ ). This differentiation was mainly due to the fast evolution at *VRN-1B<sub>int1</sub>* ( $\theta = 0.168$  over all generations,  $P = 0.0002$ ). For each population, pairwise tests of temporal differentiation showed that evolution between generations 7 and 12 was never significant whatever the locus.

**Spatial differentiation.** Divergent evolution of the DM populations for the three copies of the *VRN-1* gene appeared since the second generation and increased significantly until G12 ( $\theta_{G2} = 0.02$ ,  $\theta_{G7} = 0.15$  and  $\theta_{G12} = 0.27$ ;  $P = 0.0002$  for the three values). Pairwise tests of differentiation showed that the population of Le Moulon was the most differentiated population at generations 7 and 12.

**Evolution of linkage disequilibrium and multilocus haplotype reconstruction.** Strong linkage disequilibrium (LD) was found between the three polymorphic markers (*VRN-1A<sub>pr</sub>*, *VRN-1A<sub>ex7</sub>* and *VRN-1B<sub>int1</sub>*) within the parental lines and

between the two polymorphic loci of *VRN-1A* in all the populations (Fig. 1). LD decreased between the parents and G0 and correlations between pairs of loci were lower in G2 than in the parents. However, in two populations out of three, associations between the different polymorphic loci increased across generations even between different copies of the *VRN-1* gene, involving polymorphic loci from genomes A and B for Vervins, and polymorphic loci from the three genomes for Toulouse (Fig. 1).

Given the level of LD observed, genetic diversity and spatio-temporal differentiation were also considered at the multilocus level. Multilocus haplotype reconstruction revealed 15 haplotypes overall (Table 1). Three haplotypes were present in the parental population (*h1* and *h8*, spring and *h3*, winter; Fig. 2). New haplotypes appeared since the first generation G0 and at the following generations in the different populations possibly due to recombination (*h5*, *h6* or *h9* for example), or due to the emergence of alleles not present in the parental population (*h4* and *h11* to *h15*). Only two haplotypes (both of which were parental) were found in the Le Moulon population at G12. In contrast, seven and six haplotypes were found in the populations of Vervins and Toulouse, respectively, among which, one and three haplotypes were already found within the parents, respectively. In these populations, some of the new haplotypes strongly increased in frequency. In the Vervins population, haplotypes with a new allele at *VRN-1A<sub>pr</sub>* were at a high frequency (cumulated frequency of 0.22) since generation 2 and increase up to 0.68 in generation 12 (see for example *h15*). In the Toulouse population, the new haplotype *h5* coming from the recombination of the most frequent haplotype *h3* with a haplotype carrying a spring allele at *VRN-1B<sub>int1</sub>* increased continuously from G2 (0.04) to G12 (0.25). Estimated haplotype diversity within population evolved in the same way as genotypic diversity averaged over the four loci. Whereas diversity decreased in Le Moulon (from 0.46 in G2 to 0.22 in G12), it increased in Toulouse and Vervins between G2 and G7 and remained stable between G7 and G12 (from 0.53 in G2 to 0.74 in G12 in Toulouse and from 0.46 to 0.74 in Vervins).

Spatial differentiation estimated on haplotypic data significantly increased across generations (Table 2a). AMOVA performed on *VRN-1* haplotypes showed that 10% of the variation was ascribed to differences among populations, whereas 7% was ascribed to generations within populations (Table 3). This suggests that spatial effects of *VRN-1* haplotype differentiation are stronger than temporal effects.

#### *Spatio-temporal evolution of growth habit and heading date*

Exact tests of differentiation revealed no significant evolution of growth habit frequencies between the parents and the initial DM population G0. In the Le Moulon

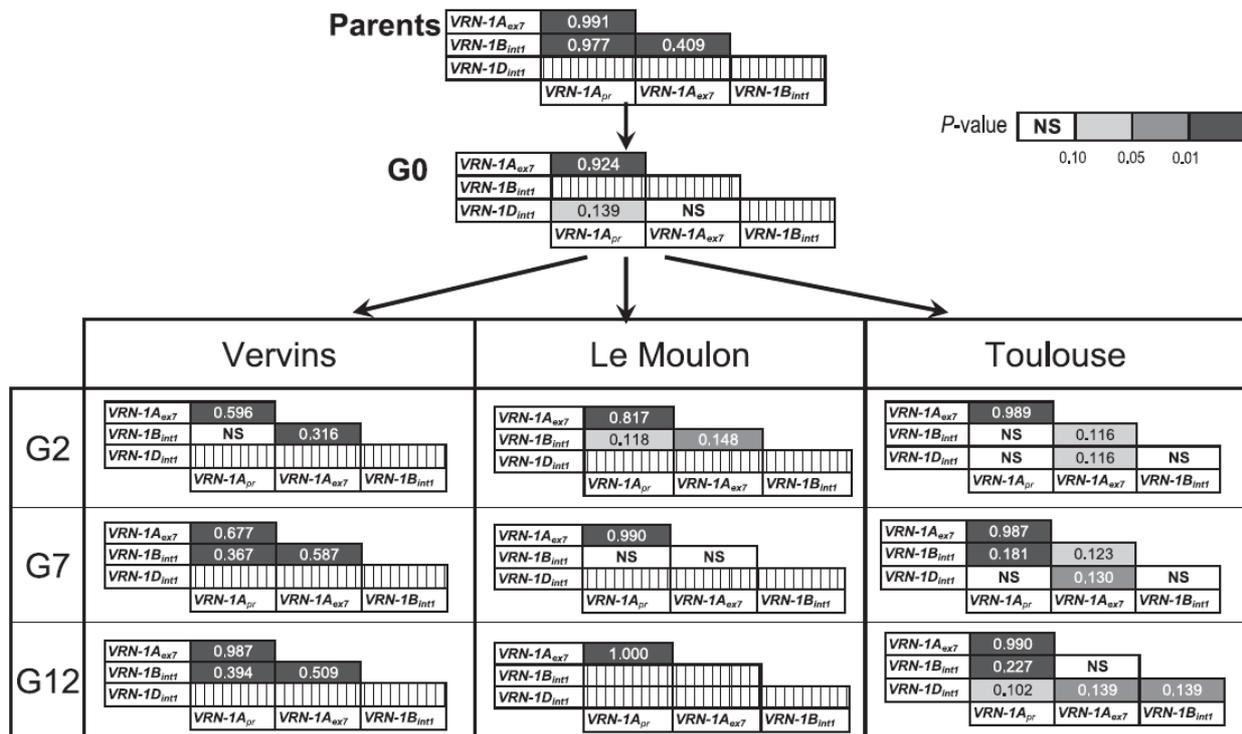


Fig. 1 Significant linkage disequilibrium (LD) between pairs of *VRN-1* loci in the parental population, the initial population G0 and in the three DM population evolved in Le Moulon, Toulouse and Vervins sampled at generations G2, G7 and G12. Nonsignificant (NS) LD values are not reported. Striped boxes indicate inestimable correlations due to the fixation of an allele at one of the two loci.

**Table 1** Haplotypes reconstructed from the four polymorphic *VRN-1* loci and association with the growth habit in the DM populations. The column ‘Mean-HdNV associated’ indicates the mean value of heading date without vernalization (HdNV) estimated on the homozygous genotypes having the corresponding haplotype at the *VRN-1* gene. The column ‘Growth Habit associated’ indicates the growth habit associated with the corresponding haplotype. *h1* to *h10*: haplotypes with an allele described in Yan *et al.* (2004) at *VRN-1A<sub>pr</sub>*, *h11* to *h15*: haplotype with an allele not previously described in (Yan *et al.* 2004) at *VRN-1A<sub>pr</sub>*

Haplotypes	<i>VRN-1A<sub>pr</sub></i> (5 alleles)	<i>VRN-1A<sub>ex7</sub></i> (2 alleles)	<i>VRN-1B<sub>int1</sub></i> (2 alleles)	<i>VRN-1D<sub>int1</sub></i> (2 alleles)	No. of individuals with this haplotype	Mean-HdNV associated	Growth habit associated
<i>h1</i>	VRN (1)	VRN (2)	VRN (2)	vrn (1)	6	932,1	S
<i>h2</i>	VRN (1)	VRN (2)	vrn (1)	vrn (1)	9	1029,9	S
<i>h3</i>	vrn (2)	vrn (1)	vrn (1)	vrn (1)	301,5	1893,7	W
<i>h4</i>	vrn (2)	vrn (1)	vrn (1)	VRN (2)	8	949,9	S
<i>h5</i>	vrn (2)	vrn (1)	VRN (2)	vrn (1)	28	1135,3	S
<i>h6</i>	vrn (2)	VRN (2)	vrn (1)	vrn (1)	0,5	NA	NA
<i>h7</i>	vrn (2)	VRN (2)	VRN (2)	vrn (1)	2	1389,2	S
<i>h8</i>	VRN (5)	VRN (2)	vrn (1)	vrn (1)	67,5	1044,1	S
<i>h9</i>	VRN (5)	vrn (1)	vrn (1)	vrn (1)	2	1101,3	S
<i>h10</i>	VRN (5)	VRN (2)	VRN (2)	vrn (1)	4,5	935,8	S
<i>h11</i>	vrn (3)	vrn (1)	vrn (1)	vrn (1)	7	1908,6	W
<i>h12</i>	vrn (3)	vrn (1)	VRN (2)	vrn (1)	6	1340,9	S
<i>h13</i>	vrn (4)	vrn (1)	vrn (1)	vrn (1)	2	995,4	S
<i>h14</i>	vrn (4)	VRN (2)	vrn (1)	vrn (1)	23,5	1950,2	W
<i>h15</i>	vrn (4)	VRN (2)	VRN (2)	vrn (1)	38,5	1332,5	S

vrn, allele associated with winter type; VRN: allele associated with spring growth habit (according to Sherman *et al.* 2004; Yan *et al.* 2004; Fu *et al.* 2005; and the present study). 1,2,3,4,5 refer to the allele code number attributed in the article. S, spring growth habit; W, winter growth habit. NA: non available.

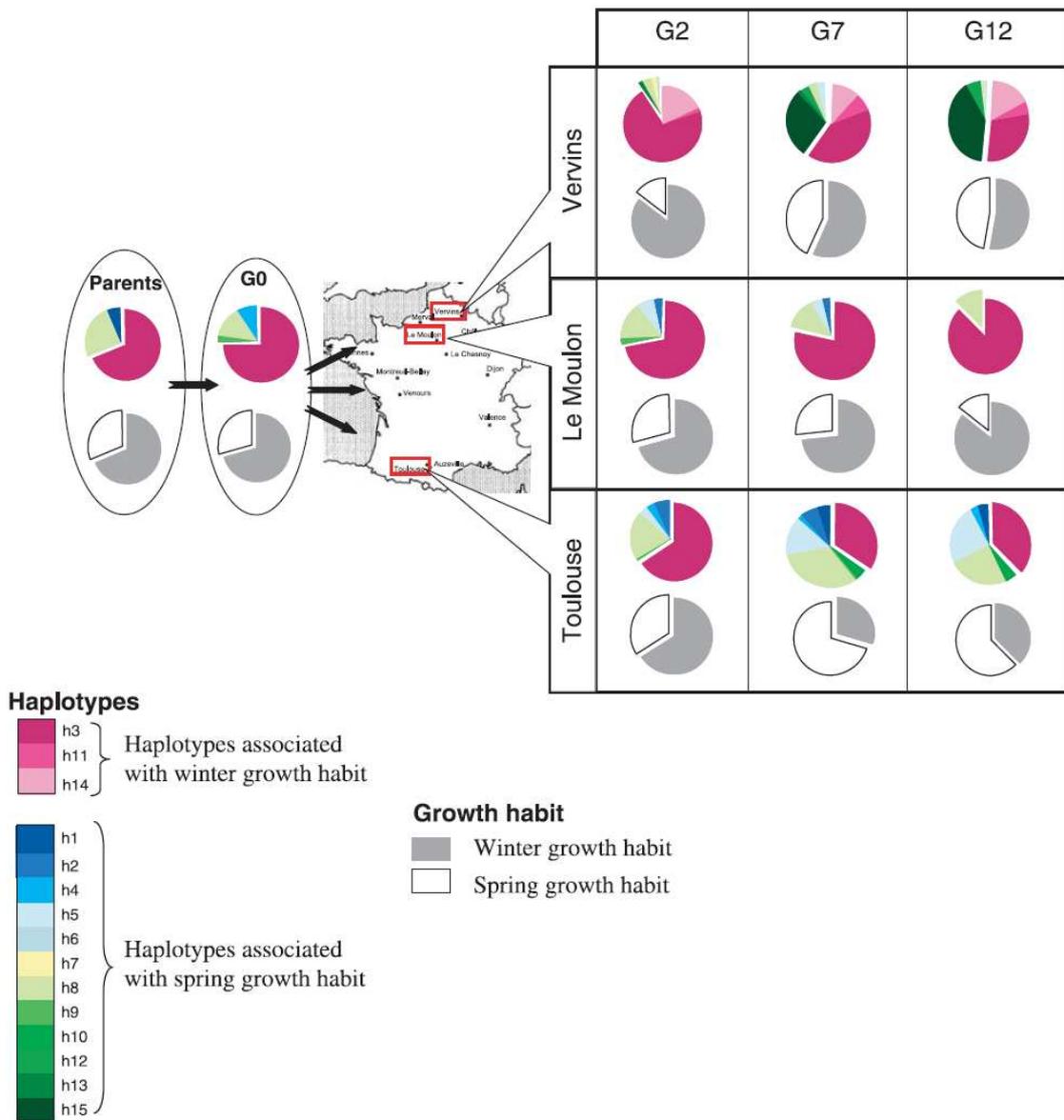


Fig. 2 Haplotypes (coloured pie chart) and growth habit (black and white pie chart) frequencies in the parental population, in the initial population G0 and in the three DM populations (Le Moulon, Toulouse and Vervins) across generations G2, G7 and G12. Haplotypes associated with spring growth habit are grouped together (as for haplotypes associated with winter growth habit).

population, the frequency of winter growth habit tended to increase from 0.71 in G2 to 0.86 in G12 (evolution significant between G7 and G12,  $P = 0.04$ ). In contrast, the frequency of spring growth habit significantly increased in the Toulouse population to reach 0.66 in G12. In the population of Vervins, the growth habit frequencies evolved in a more erratic way. Whereas the winter growth habit frequency significantly ( $P = 0.02$ ) increased between G0 (0.71) and G2 (0.85), spring growth habit frequency strongly increased between G2 and G7 (from 0.15 to 0.43;  $P < 0.0001$ ) to remain stable until G12 (0.47). Temporal differentiation

for growth habit, taking into account G2, G7 and G12, was nonsignificant in the Le Moulon population ( $\theta = 0.015$ ,  $P = 0.15$ ) but was significant and equal to 0.122 ( $P = 0.0006$ ) and 0.116 ( $P = 0.0006$ ) in the Toulouse population and the Vervins population, respectively. Spatial differentiation increased continuously across generations from 0.034 in G2 (nonsignificant) to 0.206 ( $P = 0.0002$ ) in G12.

Analysis of variance performed with the model (1) showed a significant effect of the population ( $F = 32.17$ ;  $P < 0.0001$ ) on the heading date of nonvernalized genotypes (HdNV; Table 4). The two populations from the

**Table 2a** Spatial differentiation estimated on the variance of *VRN-1* haplotype frequencies ( $\theta_{Hapl}$ ) and on the growth habit frequencies ( $\theta_{GH}$ ) among all populations at each generations

	All populations	
	$\theta_{Hapl}$ (P value)	$\theta_{GH}$ (P value)
G2	0.03 (< 0.001)	0.03 (< 0.067)
G7	0.15 (< 0.001)	0.17 (< 0.001)
G12	0.25 (< 0.001)	0.21 (< 0.001)

**Table 2b** Temporal differentiation estimated on the variance of *VRN-1* haplotype frequencies.

	All generations (G2-G7-G12)	
	$\theta_{Hapl}$ (P value)	$\theta_{GH}$ (P value)
Vervins	0.12 (< 0.001)	0.12 (< 0.001)
Le Moulon	0.01 (0.099)	0.01 (0.151)
Toulouse	0.05 (< 0.001)	0.12 (< 0.001)

north (Le Moulon and Vervins) had higher values of HdNV ( $LSMeans_{LM} = 1744$ , +95°days compared to the initial population;  $LSMeans_{VV} = 1719$ , +70°days,  $LSMeans$  estimated considering the three studied generations for a given site) than the Toulouse population from the South ( $LSMeans_{TO} = 1432$ , -87°days). A lower but significant effect of generation within population was also detected ( $F = 5.28$ ;  $P < 0.0001$ ). Globally, model (1) explained 18% of the HdNV variation.

#### Relations between *VRN-1* allelic variations at the three genomes and earliness without vernalization

The effects of the three copies of *VRN-1* gene on HdNV variation (model 2) are highly significant, and explain all together 80% of the phenotypic variation (Table 4). The interaction between *VRN-1A* and *VRN-1B* is highly significant due to the epistatic interaction between the *VRN-1* loci created by the dominance over the other loci of the spring form at one of the loci. The  $F$  value for the *VRN-1D* effect is very high as compared to other loci due to the fact that all genotypes carrying a spring allele at this locus have a low value of HdNV ( $HdNV-Means = 961$ ). Partial  $R^2$  estimated from the model 2' revealed that the

*VRN-1A*, *VRN-1B* and *VRN-1D* variations explain 51%, 23% and 5% of the total HdNV variability, respectively.

$HdNV-Means$  estimated for the winter and spring alleles at the *VRN-1B* and *VRN-1D* loci are in agreement with those found in the literature ( $vrn1B_{int1-1}$ :  $HdNV-Means = 1727$ , associated with winter type;  $Vrn1B_{int1-2}$ :  $HdNV-Means = 1225$ , associated with spring type;  $vrn1D_{int1-1}$ :  $HdNV-Means = 1642$ , winter type;  $Vrn1D_{int1-2}$ :  $HdNV-Means = 961$ , spring type). Heterozygous genotypes show intermediate  $HdNV-means$  value of earliness. The 'new' alleles found at *VRN-1A<sub>pr</sub>* are associated with winter growth habit ( $vrn1A_{pr-3}$ :  $HdNV-Means = 1646$ ;  $vrn1A_{pr-4}$ :  $HdNV-Means = 1543$ ), whereas  $HdNV-Means$  for the three other alleles were in agreement with previously published growth habit type (spring  $Vrn1A_{pr-1}$ :  $HdNV-Means = 994$ ; winter  $vrn1A_{pr-2}$ :  $HdNV-Means = 1807$ ; spring  $Vrn1A_{pr-5}$ :  $HdNV-Means = 1046$ ).  $HdNV-Means$  per haplotype are given in Table 1. Among the 10 haplotypes showing an allele previously described in Yan *et al.* (2004) at *VRN-1A<sub>pr</sub>* ( $h1$  to  $h10$ ), only one is associated to a winter growth habit ( $h3$ , Table 1). This haplotype has alleles described as 'winter allele' at each of the four loci of *VRN-1*. The other haplotypes were associated with a spring type and presented at least one 'spring' allele at one of the four *VRN-1* loci. Among the five haplotypes with a 'new' allele at *VRN-1A<sub>pr</sub>* ( $h11$  to  $h15$ ), two haplotypes are found to be associated with a winter growth habit ( $h11$  and  $h14$ , Table 1). Unexpectedly, the haplotype  $h14$  presented a 'spring allele' at *VRN-1A<sub>ex7</sub>*. Conversely, the haplotype  $h13$  showing 'winter alleles' at the four loci is associated with a spring growth habit.

## Discussion

#### Origins of alleles not present in the parental lines and found in the populations of Vervins and Toulouse

Two alleles at *VRN-1A<sub>pr</sub>* locus (previously undescribed in the literature) and the deletion in the first intron of *VRN-1D* (described in Fu *et al.* 2005) were not present within the 16 parental lines but appeared in the Vervins and the Toulouse populations, respectively. The two new alleles were also encountered in a collection of 239 hexaploid wheat lines representing diverse geographical origins and status (unpublished data), providing evidence that their presence is not an artefact specific to the DM material. Alleles not present in the parental population can appear by migration or by mutation. Migration may occur by

Source of variation	d.f.	Sum of squares	% variation	P
Between populations	2	29.794	10.03	< 0.0001
Between generations within populations	6	17.948	6.86	< 0.0001
Between individuals within generations	1059	293.379	83.11	0.001

**Table 3** Results of hierarchical AMOVA conducted on *VRN-1* haplotype frequency variation among DM populations and among generations within population. Nonparametric tests of significance were based on 5000 permutations

**Table 4** Results of ANOVA testing †the population (Le Moulon, Toulouse, Vervins) and generation (G2, G7, G12) effects on the heading date without vernalization (HdNV); ‡the effects of the three *VRN-1* copies together considering the three populations at three generations

	Model	Trait	R <sup>2</sup>	Source of variation	d.f.	Sum of squares	Mean square	F	P
†	(1)	HdNV	0.18	Population	2	900846	4504023	32.17	< 0.0001
				Generation (population)	6	4436226	739371	5.28	< 0.0001
‡	(2')	HdNV	0.78	<i>VRN-1A</i>	11	39158004	3559818	92.47	< 0.0001
				<i>VRN-1B</i>	2	17905061	8952530	232.55	< 0.0001
				<i>VRN-1D</i>	1	4157162	4157162	107.99	< 0.0001

seed contamination during harvest. Although the DM populations were grown in isolated conditions separated by at least 50 m from other fields of cultivated wheat, contamination by exogenous pollen is also still a possibility. In a field experiment on winter wheat cultivars, Hanson *et al.* (2005) found that the maximum distance of pollen-mediated gene flow was 42 m from the source. However, the potential for pollen-mediated gene flow, as well as male sterility, may differ depending on the wheat varieties and environmental conditions. In particular, high humidity and low temperature, as in Vervins, favour outcrossing in wheat. However, the new alleles found at *VRN-1A<sub>pr</sub>* were mainly found in old landraces (in the 1930s for *vrn1A<sub>pr</sub>-4*) or exotic varieties in the collection of 239 hexaploid wheat, suggesting that such varieties were not likely to be cultivated near the DM population of Vervins. Therefore, mutation should also be considered. We checked the presence of *vrn1A<sub>pr</sub>-3* and *vrn1A<sub>pr</sub>-4* in wild (*Triticum dicoccoides*), primary domesticated (*Triticum dicoccum*), and cultivated *T. durum*. Whereas *vrn1A<sub>pr</sub>-2*, *Vrn1A<sub>pr</sub>-1* and *Vrn1A<sub>pr</sub>-5* were observed in the three taxa, *vrn1A<sub>pr</sub>-3* and *vrn1A<sub>pr</sub>-4* were only amplified in cultivated bread wheat, suggesting a recent and specific appearance of these alleles (data not shown). Conversely, five new alleles amplified and sequenced on tetraploid wheat were not found in the *Triticum aestivum* species, suggesting a high mutation potential of the promoter region in the genus *Triticum*. To our knowledge, little is known on the evolutionary mechanisms modelling genetic diversity in promoters or other regulatory regions such as introns. Recurrent independent mutation events occurring in both DM and collection material could explain our results. In this study, new alleles at *VRN-1D<sub>int1</sub>* and *VRN-1A<sub>pr</sub>* appeared during the first generations of population evolution. The constitution of the initial population G0 by four generations of pyramidal crosses between the parental lines followed by three generations of multiplication is likely to produce a high level of heterozygosity. As suggested by Raquin *et al.* (in preparation), by enhancing the probability of mutation by unequal crossing-over, this phenomenon could increase the mutation rate of repeated DNA regions, such as regions containing microsatellites for example.

#### *Evolution of the growth habit frequencies and earliness in response to environmental conditions*

The populations grown in contrasted environments showed a differentiation for the growth habit qualitative trait of  $\theta = 0.21$  after 12 generations of evolution. This value is high compared to the differentiation observed for 30 neutral RFLP loci ( $\theta = 0.04$ ) previously estimated at generation 10 by Enjalbert *et al.* (1999). Estimating the expected level of differentiation for two divergent populations with a demographic size of 2650 individuals (5000 plants per populations corrected for the mean inbreeding coefficient  $\alpha = 0.9$ , Goldringer *et al.* 2001) evolving under genetic drift only during 12 generations gave also a smaller value of genetic differentiation ( $\theta = 1 - e^{-t/2N} = 0.002$ ; Hedrick 2005). This suggests that growth habit is submitted to diversifying selection. As expected, spring type frequency increased in the population from the South (Toulouse), whereas winter type frequency increased in the Le Moulon population located in the North of France. Correlatively, plants from Toulouse and from Le Moulon headed earlier and later than the initial population, respectively. This divergent evolution of winter frost requirement between populations could be interpreted as the result of adaptation of the length of the growing season to local climatic conditions. Indeed, in the South of France, winters are warmer and the vernalization requirement could be unsatisfied for winter genotypes that flowered later and may be exposed to heat and drought stresses in early summer. In contrast, in the North where the winter season is long and cold, winter growth habit genotypes waiting for the warmer season to initiate flowering and accumulating resources for seeds production are promoted, whereas spring growth habit genotypes that produced flowers too early may be exposed to frost injury.

A high frequency of spring growth habit (0.47 in G12) was however, found in the most northern population of Vervins. This result diverges from the observations of Goldringer *et al.* (2006) at generation 10, who found the highest frequency of winter genotypes (0.85 in G10) and the highest mean value of HdNV in this site. However, in Goldringer *et al.* (2006), the population of Vervins considered

was grown under an 'intensive' (conventional) farming method, whereas in this study, all the populations considered were grown under an 'extensive' condition using less nitrogen fertilizer and no fungicide treatments. Similar divergent evolution of growth habit frequencies between the two farming conditions was observed at Rennes (48.1°N, 1.4°W) (Goldringer *et al.* 2006). In the case of extensive farming conditions, plants are submitted to higher pathogen pressures, especially in Vervins known to be the site of the DM network with the strongest powdery mildew pressure in relation with local climatic conditions (Le Boulc'h *et al.* 1994; Paillard *et al.* 2000). Several studies on plant-pathogen interactions in *Arabidopsis thaliana* (Peters 1999; Korves & Bergelson 2003) and a recent review (Elzinga *et al.* 2007) showed that biotic interactions can modulate selection on flowering phenology. Indeed, genotypes with a long vegetative growth period might be subjected to a longer pathogen pressure period before reproduction than early flowering genotypes. Using spring genotypes only in the first variance analysis (model 1), we found that the mean heading date of Vervins was significantly higher than those from other populations ( $P < 0.0001$  and Fig. S1b, Supplementary material), suggesting a possible trade-off between vernalization requirement satisfaction and stress avoidance in this population.

#### Implications for *VRN-1* evolution in DM populations

Parallel evolution of growth habit and allelic frequencies (Table 2a, b) at the three copies of the *VRN-1* gene was observed in the populations studied ( $P = 0.001$  for a Mantel test on between population pairwise  $\theta$ -values obtained for haplotypic data and growth habit type data). As for the growth habit trait, the high level of differentiation of *VRN-1* ( $\theta = 0.25$  over all the populations in generation 12) compared to the neutral expectation suggested that the significant evolution of allelic frequencies at the *VRN-1* gene was the result of selection acting on the growth habit and leading to local adaptation. Due to drift and selection at the intrapopulation level, populations evolving in contrasted environments without migration are expected to lose part of their initial variability, whereas variability is expected to increase overall, due to genetic differentiation among populations. Aiming at conserving genetic diversity in crop species, dynamic management (DM) is based on such evolutionary mechanisms. Here, both a high level of total haplotypic diversity (mean  $H_e = 0.56$ ) and a significant population differentiation (Table 2a) was observed in G12. Unexpectedly, a high level of diversity was also found within two populations out of three (Vervins  $H_e = 0.40$  and Toulouse  $H_e = 0.38$  in G12). In isolated selfing populations, the genetic and haplotypic diversity is expected to be low due to reduced effective population sizes and reduced effective recombination

rates (Glémin *et al.* 2006). Previous studies of natural populations in *A. thaliana* (Mitchell-Olds & Schmitt 2006) and *Medicago truncatula* (Bonnin *et al.* 2001) showed that local spatial structure can maintain a high level of local diversity in populations of predominantly selfing plants. In wheat populations, the high level of polymorphism could be explained by several hypotheses: (i) positive frequency-dependent selection (where rare genotypes are favoured and common genotypes have lower fitness) which could explain the maintenance of haplotypes at low frequencies (*h11* and *h12* in Vervins, *h4* in Toulouse, Fig. 2), and/or (ii) temporal variations in selection across years. Inter-annual variation of climatic conditions could be responsible of the maintenance of polymorphism. In the population of Vervins, variable pathogenic pressures could temporarily favour early genotypes instead of the late genotypes as it would be expected under such climatic conditions. Moreover, in Vervins and Toulouse, the emergence of new alleles by mutation or migration also played a key role in the increase of genetic diversity, by allowing new allelic combinations to be selected.

Although we found that variation at the three copies of the *VRN-1* gene was correlated to the variation of earliness without vernalization over all populations ( $R^2_{part-A} = 0.51$ ,  $R^2_{part-B} = 0.23$ ,  $R^2_{part-D} = 0.05$ ), we also showed that the effects of the different copies varied in time and space. During the first generations, DM populations derived from the same initial pool G0 were distributed in contrasted environments and the populations faced dramatic changes in their climatic conditions of cultivation, leading to a rapid evolution of genes involved in variation of adaptive traits. Consequently, all the populations evolved very quickly between G0 and G7, whereas allelic frequencies evolved nonsignificantly between G7 and G12. Rapid phenotypic changes in the first generations could be interpreted as the result of selection at major genes leading to the fixation of favourable alleles in the first generations, whereas slight changes observed in the following generations could be the result of selection acting on multiple genes with minor effects and appearance of novel variants of genes involved in the quantitative trait (Barton & Keightley 2002). In our case, other genes with minor effects are potentially involved in the residual 20% of HdNV variation not explained by the *VRN-1* gene variation. Indeed, a recent QTL meta-analysis of Hanocq *et al.* (2007) showed that many regions of the wheat genome are involved in the control of heading date. Further studies of these regions in DM populations will allow testing this hypothesis. An increasing number of studies show that adaptation may act on few genes with large phenotypic effect contrary to the first predictions proposed in Fisher's infinitesimal model in which the mechanism of adaptation involved the fixation of many genes with small individual effects on the trait (reviewed in Orr 2005). The present

study is a new example showing that selection may act on few major genes resulting in major phenotypic changes.

*Epistatic interactions mediated through cis-regulation are involved in the response to selection on growth habit*

An increasing number of studies indicates that mutations in *cis*-regulatory sequences or in *cis*-acting regulatory genes may be an important target of adaptive evolution (for reviews, see Doebley & Lukens 1998; Purugganan 2000; Wray 2007). By modifying the level of gene expression, they produce novel phenotypes upon which natural selection can act. Interestingly, mutations in regulatory regions of *VRN-1* rather than mutations in coding regions were associated with adaptive phenotypic variation for earliness without vernalization in the DM populations. Yan *et al.* (2003) showed that *VRN-1* is expressed in spring wheat and in vernalized winter wheat. The expression of *VRN-1* in winter genotypes is controlled by repressors interacting with the promoter region and with the first intron regulatory sequences. Mutations altering repressor recognition sites at one of the three *VRN-1* copies lead to the expression of the gene in unvernallized genotypes (Loukoianov *et al.* 2005). This mechanism explains the epistatic interaction between the three copies and the dominance found for the spring allele in the literature (Stelmakh 1987). In our study, a spring allele at one of the three *VRN-1* copies is sufficient to confer a spring growth habit. However, two allelic combinations found in the population of Vervins were not associated with the expected growth habit (Table 1): (i) *h14* with only one 'spring' allele at *VRN-1A<sub>ex7</sub>* was associated with a winter growth habit, (ii) *h13* with four 'winter' alleles was associated with a spring growth habit in one homozygous genotype. The first result is in accordance with Sherman *et al.* (2004), who concluded that the substitution in exon 7 alone is not sufficient to explain the change between winter and spring growth habit. Moreover, the phenotype associated with this mutation seems to vary according to the genotype at the promoter, since the other haplotype where *vrn1A<sub>ex7</sub>-2* is the only 'spring' allele, *h6*, is associated with the spring type in the collection of 239 lines. Both results suggest that other regions of the gene or outside the gene may also strongly influence the trait variation. The polymorphism in the first intron of *VRN-1A* is a likely candidate, but it was not characterized in this study due to the lack of available codominant markers (Fu *et al.* 2005).

Recent experimental studies (Caicedo *et al.* 2004) and theoretical developments (Carter *et al.* 2005) showed that epistatic gene interactions may play an important role in the response to selection and thus in the genetic architecture of adaptation. In a worldwide collection of *A. thaliana*, Caicedo *et al.* (2004) found significant linkage disequilibrium (LD) between two interacting genes (*FRIGIDA* and

*FLC*) involved in the control of flowering time and located on separate chromosomes. This LD was interpreted as the result of selection on epistatic interactions. In the studied populations, we found that the epistatic interaction between *VRN-1A* and *VRN-1B* had a significant effect on HdNV variation ( $R^2_{part-AxB} = 0.03$ ). Moreover LD, previously reduced during the four generations of crosses between the parental lines leading to the initial population G0, increased significantly across generations between the two *VRN-1A* loci and the *VRN-1B<sub>int1</sub>* locus in the population of Vervins and between the three copies of the gene in the population of Toulouse (Fig. 1). In these two populations, the LD found between the copies of the *VRN-1* gene reached a much higher value (Fig. 1) than the LD estimated between 21 microsatellite markers (one per chromosome, *R* value = 0.09) in generation 17 in the population of Le Moulon (Rhoné *et al.* 2007). The particularly high LD between the *VRN-1* copies was probably generated by selection for the growth habit acting on loci presenting epistatic interactions. This led to the selection of allelic combinations at physically unlinked genes, located on distinct chromosomes.

## Conclusion

Experimental populations evolving under natural selection represent an interesting tool to study genetic bases of adaptation. Evolution of genes possibly involved in adaptive response can be followed together with the corresponding phenotypic traits. In this study, we demonstrated the potential role of a major gene and its allelic effects on flowering time in wheat. Despite local genetic drift and selection, a high level of adaptive diversity was still found after 12 generations in some populations, underlying the role of mutation or migration in generating new allelic combinations for selection. Additional study of these populations in the following generations and for other genes involved in earliness will further evaluate whether the main objective of the dynamic conservation of genetic resources is achieved.

## Acknowledgements

The authors thank Domenica Manicacci and Jean-Luc Jannink for careful reading of the manuscript. This work was financially supported by a grant from the French Bureau des Ressources Génétiques (BRG). This work is a part of B. Rhoné's PhD supported by the Conseil Régionale d'Ile de France.

## References

- Andersen CH, Jensen CS, Petersen K (2004) Similar genetic switch systems might integrate the floral inductive pathways in dicots and monocots. *Trends in Plant Science*, 9, 105–107.

- Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**, 11–21.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows pour la génétique des populations. Laboratoire Génome, Populations, Interactions: CNRS UMR 5000, Université de Montpellier II, Montpellier (France), <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>.
- Bonnin I, Ronfort J, Wozniak F, Olivier I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). *Molecular Ecology*, **10**, 1371–1383.
- Boutin-Ganache I, Raposo M, Raymond M, Deschepper C (2001) M13-tailed primers improve the readability and usability of microsatellite analyse performed with two different allelizing methods. *BioTechniques*, **31**, 24–28.
- Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences, USA*, **101**, 15670–15675.
- Carter AJ, Hermisson J, Hansen TF (2005) The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology*, **68**, 179–196.
- Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. *The Plant Cell*, **10**, 1075–1082.
- Dvorak J, Lukaszewski AJ (2000) Centromere association is an unlikely mechanism by which the wheat *Ph1* locus regulates metaphase I chromosome pairing between homoeologous chromosomes. *Chromosoma*, **109**, 410–414.
- Ehrenreich IM, Purugganan MD (2006) The molecular genetic basis of plant adaptation. *American Journal of Botany*, **93**, 953–962.
- Elzinga JA, Atlan A, Biere A, Gigord L, Weis AE, Bernasconi G (2007) Time after time: flowering phenology and biotic interactions. *Trends in Ecology & Evolution*, **8**, 432–439.
- Enjalbert J, David J (2000) Inferring recent outcrossing rates using multilocus individual heterozygosity: application to evolving wheat populations. *Genetics*, **156**, 1973–1982.
- Enjalbert J, Goldringer I, Paillard S, Brabant P (1999) Molecular markers to study genetic drift and selection in wheat populations. *Journal of Experimental Botany*, **50**, 283–290.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Fu DL, Szucs P, Yan LL *et al.* (2005) Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics*, **273**, 54–65.
- Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 3011–3019.
- Goldringer I, Enjalbert J, Raquin AL, Brabant P (2001) Strong selection in wheat populations during ten generations of dynamic management. *Genetics Selection Evolution*, **33**, S441–S463.
- Goldringer I, Prouin C, Rousset M, Galic N, Bonnin I (2006) Rapid differentiation of experimental populations of wheat for heading-time in response to local climatic conditions. *Annals of Botany*, **98**, 805–817.
- Goudet J (2002) FSTAT version 2.9.3.2. A Program to Estimate and Test Gene Diversities and Fixation Indices. Institute of Ecology, Lausanne, Switzerland, Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995).
- Goudet J, Raymond M, Demeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.
- Hanocq E, Laperche A, Jaminon O, Lainé AL, Legouis J (2007) Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. *Theoretical and Applied Genetics*, **114**, 569–584.
- Hanson BD, Mallory-Smith CA, Shafii B, Thill DC, Zemetra RS (2005) Pollen-mediated gene flow from blue aleurone wheat to other wheat cultivars. *Crop Science*, **45**, 1610–1617.
- Hedrick PW (2005) *Genetics of Populations*, 3rd edn. Jones and Bartlett Publishers, Sudbury, Massachusetts.
- Korves TM, Bergelson J (2003) A developmental response to pathogen infection in *Arabidopsis*. *Plant Physiology*, **133**, 339–347.
- Law CN, Worland AJ (1997) Genetic analysis of some flowering time and adaptive traits in wheat. *New Phytologist*, **137**, 19–38.
- Le Boulc’h V, David JL, Brabant P, De Vallavieille-Pope C (1994) Dynamic conservation of variability: responses of wheat populations to different selective forces including powdery mildew. *Genetics Selection Evolution*, **26** (Suppl. 1), 221–240.
- Loskutov IG (2001) Influence of vernalization and photoperiod to the vegetation period of wild species of oats (*Avena* spp.). *Euphytica*, **117**, 125–131.
- Loukoianov A, Liuling Y, Blechl A, Sanchez A, Dubkovsky J (2005) Regulation of VRN-1 vernalization genes in normal and transgenic polyploid wheat. *Plant Physiology*, **138**, 2364–2373.
- Mitchell-Olds T, Schmitt J (2006) Genetic mechanism and evolution significance of natural variation in *Arabidopsis*. *Nature*, **44**, 947–952.
- Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. *The Plant Cell*, **14** (Suppl.), S111–S130.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.
- Paillard S, Goldringer I, Enjalbert J, Doussinault G, Pope C, Brabant P (2000) Evolution of resistance against powdery mildew in winter wheat populations conducted under dynamic management. I. Is specific seedling resistance selected? *Theoretical and Applied Genetics*, **101**, 449–456.
- Peters AD (1999) The effects of pathogen infection and mutation of life history characteristics in *Arabidopsis thaliana*. *Journal of Evolutionary Biology*, **12**, 460–470.
- Purugganan MD (2000) The molecular population genetics of regulatory genes. *Molecular Ecology*, **9**, 1451–1461.
- Putterill J, Laurie R, Macknight R (2004) It's time to flower: the genetic control of flowering time. *Bioessays*, **2**, 363–373.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Remington DL, Purugganan MD (2003) Candidate genes, quantitative trait loci, and functional trait evolution in plants. *International Journal of Plant Sciences*, **164**, S7–S20.
- Rhoné B, Raquin AL, Goldringer I (2007) Strong linkage disequilibrium near the selected Yr17 resistance gene in a wheat experimental population. *Theoretical and Applied Genetics*, **114**, 787–802.
- Roux F, Touzet P, Cuguen J, Le Corre V (2006) How to be early flowering: an evolutionary perspective. *Trends in Plant Sciences*, **11**, 375–381.
- SAS Institute (2000) *SAS/STAT User Guide*, Version 8. SAS Publishing, Cary, North Carolina.
- Schlötterer C (2002) Towards a molecular characterization of adaptation in local populations. *Current Opinion in Genetics and Development*, **12**, 683–687.

- Sherman JD, Yan L, Talbert L, Dubcovsky J (2004) A PCR marker for growth habit in common wheat based on allelic variation at the VRN-A1 gene. *Crop Science*, **44**, 1832–1838.
- Staden R, Beal KF, Bonfield JK (1998) The STADEN package. *Computer Methods in Molecular Biology*, **132**, 115–130.
- Stelmakh AF (1987) Growth habit in common wheat (*Triticum aestivum* L. em. Thell.). *Euphytica*, **36**, 483–488.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics*, **8**, 206–216.
- Yan L, Loukouianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene VRN1. *Proceedings of the National Academy of Sciences, USA*, **100**, 6263–6268.
- Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J (2004) Allelic variation at the VRN-1 promoter region in polyploid wheat. *Theoretical and Applied Genetics*, **109**, 1677–1686.
- Yan L, Fu D, Li C *et al.* (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proceedings of the National Academy of Sciences, USA*, **103**, 19581–19586.
- Zucker M (2003) M fold web server for nucleic acid folding and hybridization prediction. *Nature Acids Research*, **31**, 3406–3415.

The authors are members of the 'dynamic management' research group at the plant genetics laboratory from Gif-sur-Yvette (France). The group uses wheat experimental populations to study the evolutionary mechanisms involved in the adaptation of populations and they develop more applied aspects on diversity conservation strategies.

## Supplementary material

The following supplementary material is available for this article:

**Fig. S1** (a) Distribution of mean heading date without vernalization (HdNV) in the parental population, the initial population and Le Moulon, Toulouse, and Vervins populations together. (b) Distribution of mean heading date without vernalization (HdNV) in the populations of Le Moulon, Toulouse and Vervins across generations.

**Fig. S2** Sequence alignment of the VRN-1A promoter amplified with genome specific primers VRN1AF/VRN1R defined in Yan *et al.* (2004). For the two fragments of allele vrn1A-3: (*hf*) = high fragment (identical to the sequence of vrn1A-2, GenBank EU127902), (*lf*) = low fragment (GenBank EU127903). Polymorphic regions are highlighted in gray. Boxed sections are regions putatively involved in the gene regulation as identified by Yan *et al.* (2003).

This material is available as part of the online article from:

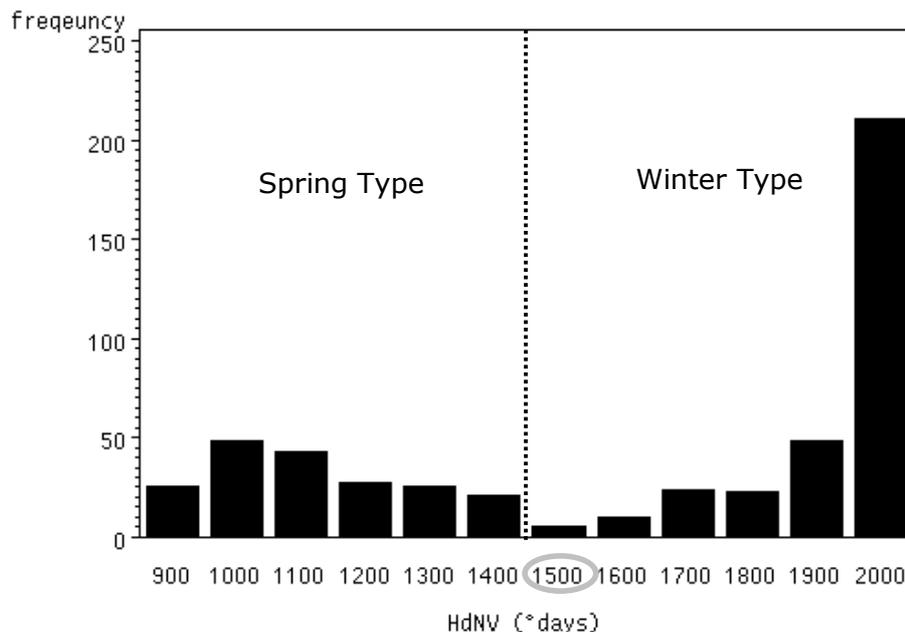
<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03619.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

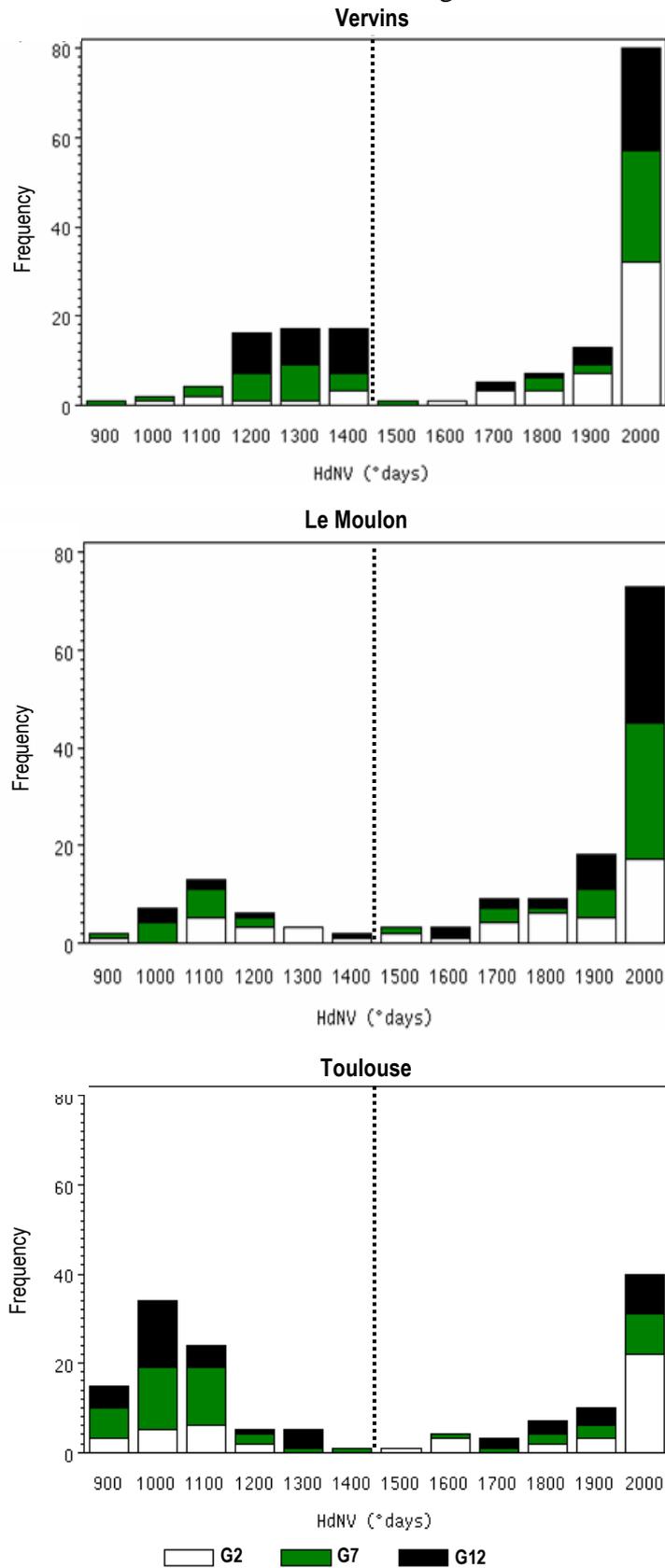
## Supplementary material:

Figure 1a: Distribution of mean heading date without vernalization (HdNV) in the parental population, the initial population and Le Moulon, Toulouse, and Vervins populations together.



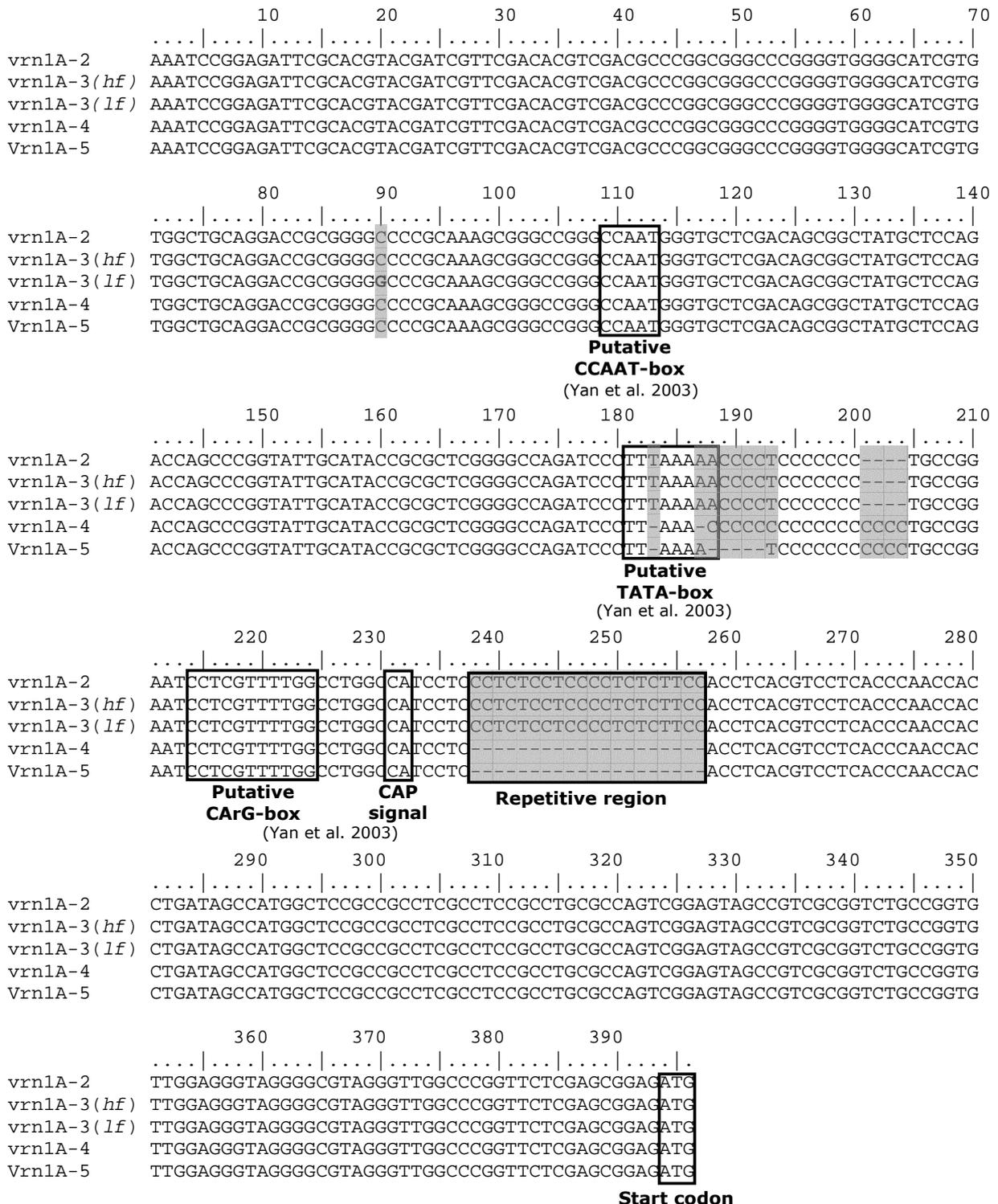
**Supplementary material:**

Figure 1b: Distribution of mean heading date without vernalization (HdNV) in the populations of Le Moulon, Toulouse and Vervins across generations



**Supplementary material :**

Figure 2: Sequence alignment of the *VRN-1A* promoter amplified with genome specific primers VRN1AF/VRN1R defined in Yan et al. (2004). For the two fragments of allele *vrn1A-3*: (*hf*) = high fragment (identical to the sequence of *vrn1A-2*, GenBank EU127902), (*lf*) = low fragment (GenBank EU127903). Polymorphic regions are highlighted in gray. Boxed sections are regions putatively involved in the gene regulation as identified by Yan et al. (2003).

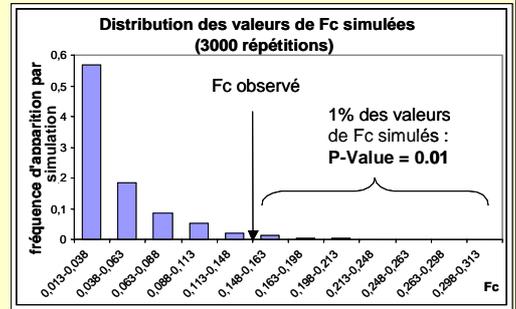


Encadré V-1 : Présentation des méthodes de détection de la sélection utilisées dans cette étude  
(N.B. : L'objectif ici est de présenter les méthodes telles qu'elles ont été développées dans le cas général. Dans l'article, ces méthodes ont été modifiées en tenant compte des paramètres démographiques connus des populations).

**Méthode Goldringer et Bataillon (2004)**

➔ **Détection de la sélection temporelle d'un locus  $L$  dans une population**

- 1- Génotypage d'un grand nombre de marqueurs pour une même population échantillonnée à 2 générations différentes
- 2- Paramètres de populations estimés à partir des marqueurs : **variance temporelle de fréquences alléliques ( $F_c$ )** entre 2 générations et **Effectif efficace ( $N_e$ )** à partir du  $F_c$  moyen
- 3- Obtention par simulation de la distribution des  $F_c$  pour une population évoluant sous dérive pure, de même valeur de  $N_e$  qu'estimée précédemment sur l'ensemble des locus et présentant les mêmes fréquences alléliques que celles observées au locus  $L$  dans la population initiale considérée.
- 4- La valeur de  $F_c$  observée au locus  $L$  est confrontée à la distribution de  $F_c$  simulés

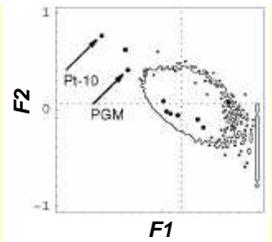


**Méthode Vitalis et al. (2001)**

➔ **Détection de locus sous sélection divergente entre populations 2 à 2**

- 1- Génotypage d'un grand nombre de locus dans les différentes populations considérées
- 2- Paramètres de populations estimés à partir des marqueurs : **statistiques  $F_i$  et  $F_j$**  (paramètre reflétant la divergence par rapport à l'ancêtre commun) à chaque locus pour chaque paires de populations  $i$  et  $j$ , et **statistiques  $F_i$  et  $F_j$  multilocus** (moyenne pondérée par l'hétérozygotie à chaque locus)
- 3- Obtention par simulation de la distribution conjointe des  $F_i$  et  $F_j$  pour des populations présentant une origine commune et évoluant sous **divergence pure** (dérive pure, pas de migration) en tenant compte des paramètres de nuisances (effectifs efficaces de la population ancestrale et taux de mutations)
- 4- Les valeurs des paramètres  $F_i$  et  $F_j$  estimés à chaque locus sont confrontées à la distribution conjointe des  $F_i$  et  $F_j$  simulés conditionnellement au nombre d'allèles au locus considéré

Valeurs de  $F_1$  et  $F_2$  estimées entre 2 populations de drosophiles pour 43 allozymes. La ligne représente l'enveloppe contenant 95% de la distribution des valeurs de  $F_1$  et  $F_2$  obtenue par simulation de populations évoluant en divergence pure pour un nombre d'allèles  $k=4$ . Les flèches pointent les locus sortant de la distribution

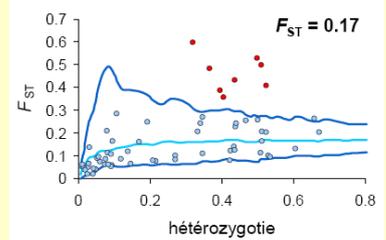


**Méthode Beaumont et Nichols (1996)**

➔ **Détection de locus sous sélection divergente entre plusieurs populations**

- 1- Génotypage d'un grand nombre de locus dans les différentes populations considérées
- 2- Paramètres de populations estimés à partir des marqueurs : **différenciation ( $F_{st}$ )** à chaque locus et  **$F_{st}$  multilocus** (moyenne pondéré par l'hétérozygotie de chaque locus).
- 3- Obtention par simulation de la distribution des  $F_{st}$  en fonction de l'hétérozygotie pour des populations évoluant selon un **modèle en îles** et présentant le même niveau de différenciation que le  $F_{st}$  moyen multilocus estimé sur l'ensemble des marqueurs.
- 4- Les  $F_{st}$  estimés à chaque locus sont confrontés à la distribution simulée en fonction de leur hétérozygotie.

Valeurs de  $F_{st}$  estimées pour 61 allozymes dans 15 populations de drosophile en fonction de l'hétérozygotie (Beaumont et Nichols 1996). Les lignes représentent les quantiles à 0.025, 0.5 et 0.925 de la distribution obtenue par simulation de populations évoluant sous un modèle en île avec  $F_{st}=0.17$



## V-3 Détection de la sélection des gènes candidats pour la précocité de floraison

Dans cet article en préparation, l'évolution de *VRN-1* ainsi que de 5 autres gènes potentiellement impliqués dans le caractère de précocité est mise en relation avec la précocité estimée au champ sur le site du Moulon. Parmi les 5 autres gènes étudiés, 2 (*Ppd* et *FT*) publiés dans la littérature, sont connus comme étant impliqués dans l'architecture du caractère : *Ppd* intervient dans la voie de la photopériode tandis que *FT* est un gène intégrateur des voies de la photopériode et de la vernalisation. Pour les 3 autres gènes (*LD*, *GI* et *CO*) non encore publiés chez le blé, des polymorphismes nucléotidiques ont été mis en évidence chez les 16 parents des populations. Nous avons tout d'abord démontré indirectement que la précocité au champ est sélectionnée entre les populations, *via* l'estimation de leur divergence pour ce trait ( $Q_{ST}$  comparé au  $F_{ST}$  de marqueurs neutres). Outre la réalisation d'une étude d'association entre les polymorphismes des 6 gènes et la variation de précocité dans les 3 populations de gestion dynamique, nous avons aussi utilisé 3 méthodes différentes de détection de locus sélectionnés (Encadré V-1). Les paramètres tels que les fréquences alléliques initiales, les tailles efficaces des populations, et le temps de divergence, qui sont inconnus dans les populations naturelles et sont des sources de nuisance dans les simulations, ont pu ici être pris en considération.

L'étude d'association entre la précocité et les 14 zones polymorphes étudiées individuellement ou par gène sous forme de combinaisons alléliques entre copies (haplotypes), montre que les gènes sont plus ou moins impliqués dans la variation du caractère, certains présentant des associations à travers leur forme haplotypique ou à travers leurs combinaisons haplotypiques avec d'autres gènes. Parmi les gènes considérés, *VRN-1*, fortement associé au caractère, est le gène montrant les traces de sélection les plus évidentes entre populations en génération 12 notamment pour l'haplotype, la sélection divergente aux locus pris individuellement étant cachée par les interactions entre gènes, en particulier pour *VRN-1B*. Pour le polymorphisme au gène *CO* présentant également une forte association avec le caractère mais uniquement dans la population Vervins, la sélection observée semble être un effet indirect de la sélection agissant sur le gène *VRN-1* avec lequel il est en déséquilibre de liaison bien que les 2 gènes soient situés sur des bras différents du même chromosome. Cette observation est confortée par les forts déséquilibres de liaison observés dans cette population

entre les locus microsatellites pourtant tous localisés sur des chromosomes différents. Enfin parmi les gènes ne présentant pas de traces de sélection malgré leur implication possible dans la variation du caractère (notamment *LD* le seul parmi les gènes non publiés dont la variation est associée à la précocité en analyse de génétique d'association sur collection (*cf.* Chapitre I) ou *Ppd* parmi les gènes publiés), deux facteurs tendent à expliquer cette observation : (*i*) soit les méthodes utilisées ne permettent pas de détecter la sélection car pour ces locus, la sélection agit majoritairement en favorisant des combinaisons alléliques particulières générant de faibles variations de fréquences alléliques à chaque locus ; (*ii*) soit les locus participent à l'architecture d'autres caractères (pléiotropie) ou sont en déséquilibre de liaison avec d'autres gènes liés à la valeur sélective et l'absence de signal sélectif est expliquée par des pressions de sélection antagonistes agissant sur ces gènes. Cette étude met en évidence qu'un certain nombre de locus participant à l'architecture des caractères polygéniques adaptatifs pourraient ne pas être détectés avec les méthodes classiquement utilisées pour détecter la sélection de locus en populations naturelles.

**Evolution of flowering time in experimental wheat populations:  
a comprehensive approach to detect genetic signatures of natural  
selection.**

**Bénédicte Rhoné, Renaud Vitalis\*, Isabelle Goldringer, Isabelle Bonnin**

Manuscrit en préparation

Station de Génétique Végétale, INRA-CNRS-INAPG-UPS, Ferme du Moulon, 91190 Gif sur Yvette, France. Tel: +0033 (0)1 69 33 23 44; Fax: +0033 (0)1 69 33 23 40

\* Unité Éco-Anthropologie et Ethnobiologie, UMR CNRS 5145 - MNHN USM 104, Équipe "Génétique des Populations Humaines", Musée de l'Homme, 17 place du Trocadéro, 75116 Paris, France.

Corresponding author: B. Rhoné

Station de Génétique Végétale, INRA-CNRS-INAPG-UPS, Ferme du moulon, 91190 Gif sur Yvette, France. Tel: +0033 (0)1 69 33 23 44; Fax: +0033 (0)1 69 33 23 40

**Running title:** Evolution of flowering time in wheat populations

## **Abstract**

In annual plant species, flowering time is a major adaptive trait that synchronizes the initiation of reproduction with favorable environmental conditions. In order to understand how evolved diversity and spatial structure of genes involved in the control of an adaptive trait submitted to selection, we studied the evolution of flowering time, genes associated with this trait and microsatellite markers supposed to be neutral in three experimental wheat populations grown in contrasting environments (Northern to Southern France). By comparing phenotypic to presumably neutral variation, we showed that flowering time was selected during the 12 generations of evolution. We then tested whether six candidate genes, presumably involved in the trait expression, exhibited signatures of selection that could be identified by spatial or temporal methods adapted for our experimental population design using three different methods of detection. Only genes with a major impact on flowering time were detected as responding to selection. With low-to-moderate phenotypic effect, or when covariance among genes occurs, even combining several spatial methods of outlier loci detection appeared inefficient. In such a case, it seems necessary to consider multilocus allelic combinations between genes, which could be targets of selection.

**Keywords:** Adaptation, Differentiation, Diversifying selection, Heading date, *Triticum aestivum* L.

## Introduction

Local adaptation resulting from spatial and temporal heterogeneity in selection pressures acting on heritable traits is the main force shaping phenotypic diversity in natural populations (Schlötterer 2002). Understanding the genetic mechanisms involved is therefore one of the greatest challenges of evolutionary biology. Yet the question of how the selection acting on an adaptive trait influences the evolution of genes involved in the structure of that trait is still largely unanswered (Orr 2005, Ehrenreich & Purugganan 2006). To get insight into the genetic mechanisms of adaptation, the identification of genes that influence ecologically important traits is a prerequisite, before detecting those responding to selection during adaptation. Flowering time is a major adaptive trait in plant species because it ensures reproduction to occur in favourable conditions with respect, e.g., to climate, herbivory, pathogen pressures, or the presence of pollinators (Remington & Purugganan 2003; Roux et al. 2006). The genetic architecture of this trait has been extensively studied in model species and crops. For example, in *Arabidopsis thaliana*, more than 80 genes involved in the floral transition have been identified (Putterill et al. 2004) revealing that flowering time is a complex trait governed by many interacting genes.

In the present work, we studied the evolution of flowering time and genes associated with this trait in three experimental wheat populations grown in contrasting environments, from Southern to Northern France. The three experimental populations were derived from the same initial population, and evolved independently for 12 generations. Wheat (*Triticum aestivum* L.) is an annual autogamous crop. Under temperate climates, it is usually sown in autumn and requires winter frost (vernalization) and exposure to long photoperiod before flowering. These major environmental signals ensure that the plant flowers in spring under favorable conditions (Loskutov 2001; Goldringer et al. 2006). In a preliminary study, Goldringer et al. (2006) found a latitudinal differentiation for earliness in these experimental wheat populations, after only 10 generations. Populations from the North exhibited both a higher frequency of the winter growth habit genotypes (which need a period of exposure to low temperature to initiate flowering) and a later flowering time than populations from the South, presumably in response to local climatic conditions. These findings led to questions about the genetic changes underlying the phenotypic differentiation among populations on such a small time scale.

We examined six candidate genes, potentially associated with the variation in flowering time. In this allo-hexaploid species (genome AABBDD), each gene is usually present in three

copies across the A, B, and D genomes. Three genes out of the six have already been identified and described in hexaploid wheat: (i) *VRN-1* is the major gene involved in the vernalization pathway. Mutations in the promoter of the genome A copy (*VRN-1A*) and in the first intron of the two genome B and D copies (*VRN-1B* and *VRN-1D*) were found to be strongly associated with the variation for flowering time without vernalization and for growth habit (spring or winter depending on the vernalization requirement ;Yan et al. 2004; Sherman et al. 2004; Fu et al. 2005; Rhoné et al. 2008). (ii) *Ppd-1* is a major gene involved in the photoperiod pathway. The three copies have recently been isolated and two mutations (a large deletion upstream of the coding region and a deletion in the last exon) have been identified in the genome D copy of the gene as involved in the variation of photoperiod sensitivity (Beales et al. 2007). (iii) An orthologous gene of *FLOWERING LOCUS T (FT)* gene of *Arabidopsis thaliana* was isolated recently in hexaploid wheat. *FT* polymorphisms (also referred to as *VRN-3* in Yan et al. 2006) are expected to affect both vernalization and photoperiod pathways as it is considered a key integrator gene (ultimately regulated by internal and external signals). In a global association study, *FT* mutations in the A and D copies were found to be mostly associated with flowering time without vernalizing treatment (Bonnin et al. 2008). The three other gene we studied were orthologous sequences of *A. thaliana* genes, known in this species to be involved either in the autonomous pathway as *LUMINI-DEPENDENS (LD)* (Lee et al. 1994) or in the photoperiod pathway as *CONSTANS (CO)* and *GIGANTEA (GI)* (Griffiths et al. 2003; Dunford et al. 2005).

Most of the approaches that have been proposed to detect marker loci potentially targeted by selection aim at detecting some loci exhibiting unusual patterns of variation compared to expectations under the neutral model (see the reviews by Beaumont 2005; Storz 2005; Vasemägi & Primmer 2005). Such approaches rely on the fact that demography affects the genome evenly, while selection for a given trait presumably affects a few genes at a time, in a locus-specific manner. Hence, regions of the genome that are driven by directional selection are expected to show a significant decline in genetic diversity, increased differentiation among populations and increased linkage disequilibrium compared to expectations under the neutral model (Vasemägi & Primmer 2005). Statistical methods based on estimates of population differentiation (as measured, e.g., by the parameter  $F_{ST}$ ) have been developed to identify singular loci that exhibit higher (or lower) population differentiation than expected under neutrality (see Beaumont & Nichols 1996; Vitalis et al. 2001; Beaumont and Balding 2004, reviewed in Beaumont 2005). A similar approach based on the temporal changes in allele frequencies (estimated as the standardized variance in allele frequencies  $F_c$ ), was

developed in order to identify loci that exhibit unusual temporal variation of allele frequencies (extreme  $F_c$ ) as compared to the rest of the genome (Goldringer & Bataillon 2004). There are many recent studies in non-model organisms that use one or several of these approaches in order to identify genomic regions targeted by selection (see, e.g., Wilding et al. 2001; Kayser et al. 2003; Rogers & Bernatchez 2005; Bonin et al. 2006; Oetjen & Reusch 2007; Bonin et al. 2007; Rhoné et al. 2007; Stinchcombe & Hoekstra 2008). In the present study, we used some of the approaches cited above to specifically test whether the candidate-gene polymorphisms, known to be associated with the variation of flowering time in the different studied populations, were targeted by selection. Simulations of the distribution of different parameters in the neutral case were obtained, based on the variation of 21 microsatellite loci genotyped in the studied populations. The use of experimental populations may increase the power of such methods, because many of the demographic parameters that are usually considered as nuisance parameters (for example the initial genetic composition, the time since divergence from an ancestral population, the rate of dispersal, and the census size of the populations) are controlled or known.

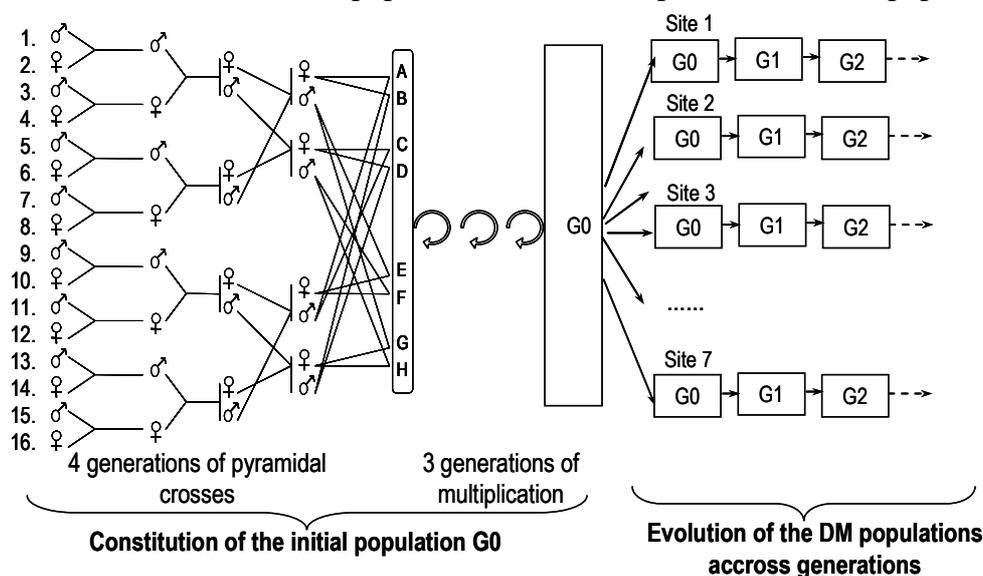
Our main objective in the present study was to investigate whether selection can be detected at candidate genes, which are potentially involved in the variation of a quantitative trait under diversifying selection. We first investigated whether flowering time was responding to natural selection in the wheat populations studied during the course of the experiment. In that aim, we compared the quantitative genetic variation for this trait ( $Q_{ST}$ ) with the genetic variation measured at 21 microsatellite loci presumed to be neutral. Then we evaluated whether the polymorphisms found at six candidate genes were associated with variation in flowering time, in order to better characterize the genetic architecture of that trait. Finally, we tested whether the candidate genes exhibited signatures of selection that could be identified by the spatial or temporal methods discussed above adapted for our experimental population design.

## Materials and methods

### Plant material and experimental design

The experimental populations studied here were part of a dynamic management program of bread wheat. Dynamic management (DM) is an evolutionary approach for genetic resource conservation, which aims to maintain genetic diversity of crop species by growing successive generations of an initial composite population in several contrasting environments, without intentional selection. These experimental populations are subject to the same evolutionary forces as are natural populations (mainly, natural selection and genetic drift). They are used as a model for studying the mechanisms involved in local adaptation to environmental conditions. The initial DM population of wheat was created by crossing 16 diverse inbred lines (with contrasting values for earliness, height, resistance levels and other fitness-related traits) for four successive generations (see Figure 1). The four consecutive crosses aimed to maximize recombination between the different parental alleles. After three generations of seed multiplication, the composite population obtained (referred to as G0) was used as the initial population of the DM program. In 1984, seed samples of G0 were sown in 7 sites of a French experimental network. Since then, each population has been cultivated every year without human selection and without migration between populations. Each year, the whole plot was harvested in bulk, a random sample of at least 5000 seeds was sown in a 100m<sup>2</sup> plot isolated from other wheat cultures to constitute the next generation while 500-1000 grams of seeds were stored in a cold room (for more details see Goldringer et al. 2006 and Rhoné et al. 2007).

Figure 1: Constitution of the initial population G0 of the experimental wheat populations



In the present study, we analyzed the spatial and temporal evolution of flowering time in the DM populations using (i) the sixteen parental lines; (ii) 36 inbred lines derived from the initial population G0 and (iii) seed samples taken at generation two (G2 grown in 1986), generation seven (G7; 1991), and generation 12 (G12; 1996) from populations grown in three distinct locations: Vervins (49.5°N, 3.5°E, northern France), Le Moulon (48.4°N, 1.4°W, near Paris) and Toulouse (43.4°N, 1.3°E, southern France). Overall, 540 individuals were sampled in the three populations studied (60 genotypes for each population at each generation). The G0 individuals used in this study consisted of inbred lines derived from individuals of the initial population with at least three generations of selfing. Seeds of the 540 individuals of the DM populations were sampled from the seed stock stored in a cold room each year during the experiment. In 2003, all the seeds in the experiment (592 seeds total) were sown at Le Moulon at low density in small plots consisting of six rows of 1.35m each. Each plant was identified as a single genotype. We took a leaf sample from each plant for DNA extraction and grew the selfed progenies of each genotype to estimate flowering time. Because wheat is a predominantly selfing species, the progenies obtained by selfing the 592 plants were highly homogenous, both genetically and phenotypically. Flowering time was measured in a common garden experiment at Le Moulon in 2004 and 2005. In early November, 20 seeds per family were sown in single-rows plots at low density in a randomized block design with two replicates in 2004 and with a single replicate in 2005. Individual plants were considered as headed when half of the main ear had emerged from the flag leaf. Because heading date is easier to assess than flowering time in wheat and since both traits are highly correlated, we recorded heading date of each family. For each plot, the heading date was recorded when half of the individuals of the row had headed. Hence, the heading date attributed to each plot is an average value. Heading dates were transformed into cumulative degree-days (i.e. the sums of the mean temperature per day) from germination to heading.

### **Molecular data**

For each plant, total DNA was extracted from 200 mg of young leaves following a protocol derived from the DNeasy 96 Plant kit of Qiagen (QIAGEN, Basel, Switzerland).

#### *Presumably neutral markers*

We chose 19 microsatellite markers (XgwmXXX) developed by Röder et al. (1998) and one primer set developed by Guyomarc'h et al. (2002) which amplified two additional loci on two

chromosomes (Cfd71a and Cfd71d), to cover the 21 wheat chromosomes. All microsatellite markers were ascertained to be polymorphic among the 16 parental lines. PCR protocols were adapted from Röder et al. (1998) and Guyomarç'h et al. (2002). PCR products were electrophoresed on an ABI 3100 semi-automated sequencer, and analyzed with the software package GeneMapper (v 3.7, Applied Biosystems).

### *Candidate-genes*

#### *VRN-1, Ppd and FT*

We characterized four polymorphic sites in the gene VRN-1: (i) the duplication, insertion and deletion in the VRN-1A promoter (denoted VRN-1Apr in the following) identified by Yan et al. (2004), (ii) a substitution in the 7th exon of VRN-1A (VRN-1Aex7) identified by Sherman et al. (2004), (iii) a 4kb deletion in the first intron of VRN-1B (VRN-1Bint1), and (iv) a 4kb deletion in the first intron of VRN-1D (VRN-1Dint1) identified by Fu et al. (2005). More details concerning the genotyping of the VRN-1 gene across the three genomes can be found in Rhoné et al. (2008).

Two polymorphic sites were characterized for the Ppd gene, both located on the D copy of the gene (Ppd-1D): a 2kb deletion upstream the coding region (denoted Ppd-1Dpr in the following) and a 16bp deletion located in the 8th exon (Ppd-1Dex8), close to a CCT domain found by Beales et al (2007). We used the same PCR conditions and digestion protocols of the PCR products as the authors'. PCR amplifications of Ppd-1Dex8, were electrophoresed on a ABI 3100 automated sequencer in order to detect variation.

Three polymorphic sites were characterized in two copies of the FT gene, across genomes A and D, identified by Bonnin et al. (2008): (i) a SNP (C/T) located in the first intron of FT-A (FT-Aint1), (ii) a variable microsatellite motif ((TG)<sub>8or9</sub>-(CG)<sub>3or4</sub>) located in the second intron of FT-A (denoted FT-Aint2a for (TG)<sub>8or9</sub> and FT-Aint2b for (CG)<sub>3or4</sub>) 130bp downstream from the SNP of the first intron, and (iii) an insertion-deletion of one G in a poly G sequence located in the third exon FT-D (FT-Dex3). These regions were sequenced following the protocol described in Bonnin et al (2008).

#### *LD, GI and CO*

From homology analyses between the sequences of *LUMINI-DEPENDENS* (*LD*), *CONSTANS* (*CO*), and *GIGANTEA* (*GI*) in rice and wheat ESTs, we designed gene-specific

primers in conserved regions of putative exons using Oligo 6 (Medprobe, Oslo, Norway). In order to discard primers giving no band or a smear, PCRs were performed on one hexaploid AABBDD accession (*Triticum aestivum* var. Recital) and three diploid accessions: *T. urartu* (AA genome), *T. speltoides* (BB genome, selfed progeny), and *T. tauschii* (DD genome). Each genomic copy amplified in the hexaploid accession was then isolated and cloned into pGEM-T Vector system I (Promega, Madison, Wisconsin, USA), and sequenced (Genome Express, Meylan, France). All sequences were analysed using the Staden software package (Staden et al. 2000). The design of genome-specific PCR primers was made possible by comparing diploid and hexaploid sequences. The genome-specificity of the pairs of primers was checked by PCR on nullitetrasonic, ditelosomic and deletion lines of the ‘Chinese Spring’ variety (Sourdille et al. 2004). Only primer pairs that amplified fragments that could be assigned to a single chromosome arm were retained for screening the 16 parental lines for polymorphisms. The PCR conditions for each pair of primers are given in the Online Supplementary Material 1. The genome-specific PCR products obtained were sequenced, and SNPs were identified from sequence alignment comparisons. All the studied populations were further genotyped for these SNPs by competitive allele-specific PCR as described in the Online Supplementary Material 1.

## Statistical analyses

### *Phenotypic analyses*

Considering spatial and temporal evolution simultaneously, analyses of variance (ANOVA) were performed on flowering time (or heading date HD) using the GLM procedure of the SAS software package (SAS Institute 2000) with the following model:

$$Y_{ijkl} = \mu + pop_i + generat(pop)_{ij} + yr_k + \varepsilon_{ijkl} \quad (1)$$

where  $pop_i$  is the effect of population  $i$  (Moulon, Toulouse or Vervins),  $generat(pop)_{ij}$  is the effect of generation  $j$  (G2, G7 or G12) nested in the population  $i$ ,  $yr_k$  is the effect of the year  $k$  of experiment (2004 or 2005) and  $\varepsilon_{ijkl}$  is the residual. All the effects were considered fixed. Because the repetition effect was not significant in 2004, we calculated the average heading date for each genotype and kept a single value for heading date per genotype and per year in the analyses. For each significant effect, adjusted means for the other effects in the model (*LS-Means*) were estimated. We tested the normal distribution of the residuals in the ANOVA with the UNIVARIATE procedure of the SAS software package. Heritability estimates were obtained using a separate ANOVA for each population at each generation with a model including random genotype effects and fixed year. Broad-sense heritabilities for each

population and generation were estimated as  $h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_E^2)$  where  $\sigma_G^2$  is the estimated among-genotype variance and  $\sigma_E^2$  is the residual variance. The variance components were estimated using the SAS VARCOMP procedure (REML method).

To estimate the spatial differentiation of phenotypes ( $Q_{ST}$ ) at each studied generation, among-populations and among-genotypes-within-population variance components were estimated using the SAS VARCOMP procedure with the following model:

$$Y_{ijk} = \mu + pop_i + geno(pop)_{ij} + yr_k + \varepsilon_{ijk} \quad (2)$$

where *geno* is the effect of genotype *j* nested in the population *i* ( $j = 1, \dots, 60$ ). Population and genotype were considered random effects. We estimated the phenotypic population differentiation ( $Q_{ST}$ ) for the flowering time trait as a ratio of among-populations and among-genotypes-within-population variance components. Because we used selfed progenies for each genotype,  $Q_{ST}$  was estimated as  $\sigma^2_{pop} / (\sigma^2_{pop} + \sigma^2_{geno})$ , which assumes complete selfing (Bonnin et al. 1996; Le Corre 2005).  $Q_{ST}$  estimates were considered significantly different from zero when the population effect was significant in the ANOVA.

### *Molecular analyses*

Allele frequencies were estimated in each population at each generation for microsatellite markers and candidate gene polymorphic sites. Genetic diversity was calculated for the two sets of markers using Nei's (1987) unbiased estimate of heterozygosity  $H_e$ . In order to detect non-random associations of alleles across loci overall the genome, we estimated the linkage disequilibrium (LD) between all pairs of microsatellite loci using the genotypic disequilibrium coefficient ( $R$ ) calculated with the LINKDIS program available in the Genetix software package (Belkhir et al. 2004). We also estimated LD between pairs of candidate gene polymorphisms. To test whether the correlations of gene frequencies across loci were significantly different from zero, we performed Fisher exact tests for gametic disequilibrium using GENEPOP version 4.0 (Raymond & Rousset 1995). We found new microsatellite alleles at low frequency at G12, which presumably appeared by mutation during the experiment. In order to reduce the effects of rare alleles on LD estimates, they were calculated only from the correlations between alleles whose frequency was higher than 5%.

### *Temporal variation of gene frequencies*

We estimated the genetic effective population sizes from the temporal variances of allele frequencies at microsatellite markers. The standardized variance ( $F_{c,l}$ ) was estimated at each locus  $l$ , between the parental lines and each of the three G12 populations as:

$$\hat{F}_{c,l} = \frac{1}{K_l} \sum_{k=1}^{K_l} \frac{(p_{ki} - p_{kf})^2}{\left( \frac{(p_{ki} + p_{kf})}{2} - p_{ki} \times p_{kf} \right)}$$

(Nei and Tajima 1981), where  $K_l$  is the total number of alleles at locus  $l$ , and  $p_{ki}$  (resp.  $p_{kf}$ ) is the frequency of allele  $k$  in the sample of the initial (resp. final) population considered. The multilocus estimate  $\hat{F}_c$  was calculated as the mean of the single locus  $\hat{F}_{c,l}$  weighted by the number of alleles at each locus. The multilocus estimate was then used to calculate the effective population size ( $N_e$ ) following Waples (1989). Confidence intervals for  $N_e$  were estimated following Goldringer & Bataillon (2004). We applied the procedure described by the latter authors to test for the homogeneity of the individual  $\hat{F}_{c,l}$  across loci and identify markers exhibiting extreme  $\hat{F}_{c,l}$  values compared to the rest of the genome. To that aim, for each of the three studied populations and for each studied gene, we simulated 3000 Wright-Fisher populations, starting with the initial allele frequencies observed in the initial population at the locus considered, and evolving for  $t$  generations of pure genetic drift ( $t = 19$  between the parental population and G12, see Figure 1), with the  $N_e$  estimated from the microsatellite markers (see above). The rationale was to use the presumably neutral markers to infer the effective population size ( $N_e$ ), and then to draw the expected distribution of  $F_{c,l}$  values under the null hypothesis of neutral evolution. The observed  $F_{c,l}$  value for each marker or candidate gene was then compared to its specific expected neutral distribution (based on its initial allele frequency) and a *p-value* was assigned to each locus. Markers or candidate gene polymorphisms that exhibit significantly larger  $F_{c,l}$  values than expected by drift alone are presumably targeted by selection. We first used this test to determine whether the microsatellite markers could be considered neutral and only those markers for which we could not reject the neutral null hypothesis were used to re-estimate  $N_e$ . This new estimate was then used in the testing procedure of each candidate gene polymorphism. Because there was evidence of epistatic interactions among the copies of the *VRN-1* gene in different genomes (Fu et al. 2005), we carried out the same tests on full haplotypes, combining the different alleles observed at each polymorphic site of a gene across the three genomes. We found new haplotypes in the final populations, which were absent from the parental lines. Those haplotypes likely appeared by recombination. Since the simulation program we used does not

account for recombination or mutation, new haplotypes were declared as already present at a very low frequency ( $f = 0.005$ ) within the parental lines when they reached a frequency higher than 5% in the final population considered.

### *Spatial variation of gene frequencies*

At any given generation, we estimated population differentiation ( $F_{ST}$ ) overall and between pairs of populations using Weir & Cockerham's (1984) estimator  $\theta$ , as calculated with the FSTAT software package (Goudet 2002). We calculated 95% confident intervals around point estimates of  $\theta$  by bootstrapping among loci (1000 samples). The significance of population differentiation was tested by means of random permutations of genotypes among samples (5000 permutations). Adjusted  $p$ -values were then obtained by applying the Bonferroni correction for multiple comparisons. We were interested in detecting loci that exhibited extreme differentiation among populations compared to that expected under neutrality. In order to detect such outlier loci, we used three distinct approaches, detailed below.

(i) We used the DETSEL method developed by Vitalis et al. (2001; 2003) that considers a pair of populations diverging from an ancestral population that may have experienced a recent change of population size (e.g., a bottleneck). This method is based on the analysis of the joint distribution of population-specific estimates of  $F$ -statistics. This approach has been shown to be only weakly dependent upon unknown parameters of the divergence model when the joint distribution is conditioned on the total number of alleles in the sample. Because our experiment consisted of three diverging daughter-populations from a single, common ancestral population, we modified the DETSEL method as to consider the simultaneous divergence of three (rather than two) populations from a common, ancestral population. Population-specific estimates of  $F$ -statistics ( $F_i$ , with  $i = 1, 2, \text{ or } 3$ ) were obtained following eq. (8) in Weir & Hill (2002). In the DETSEL approach, the multilocus estimates of the population-specific  $F_i$  parameters are used to derive the effective size of each daughter population  $N_i$ , given the divergence time  $t$ , from:  $F_i = 1 - \exp(-t / N_i)$ . Then, using the branch length estimates ( $-t / N_i$ ), as well as some values for the nuisance parameters ( $N_e$  in the ancestral population, and the mutation rate  $\mu$ ), a large number of coalescent simulations are performed, in order to derive the expected distribution under neutrality, of  $F_i$  estimates. Given our experimental setting, some of the nuisance parameters were known: the time since divergence is  $t = 12$  and the effective size in the ancestral population was likely to be of the order  $N_e = 100$ . We varied the mutation rate depending on the class of markers considered

(see below). From the simulated data, we analyzed the pairwise distributions of  $F_1$  with  $F_2$ ,  $F_1$  with  $F_3$ , and  $F_2$  with  $F_3$ , as to identify some loci exhibiting singular values of the parameters (e.g., lying out of the “high probability region” that contained, say, 95% of the simulated data points). As an addition to the original DETSEL method, we associated  $p$ -values for any given observation ( $F_1$ ,  $F_2$ ) in the real dataset. To do so, we computed the associated probability of each observation in the simulated dataset (i.e., the relative frequency of this observation over the full set of simulated data, under the null hypothesis of selective neutrality). We could then derive the empirical  $p$ -value for this observation by subtracting the sum of all cell probabilities within the simulated dataset that are more probable than the observed data from one.

(ii) Another widely used method to detect outlier loci has been proposed by Beaumont & Nichols (1996). This method is implemented in the software package FDIST2. This method is based on the analysis of the joint distribution of  $F_{ST}$  and expected heterozygosity in an island model of population structure, and the approach has been shown to be robust to violations of the model assumptions. Here, because our experimental setting better fits a pure divergence model than an island model at migration-drift equilibrium, we chose to analyze the same simulated data as generated by the modified DETSEL program above, but using the joint distribution of  $F_{ST}$  and heterozygosity instead of the  $F_i$  statistics. This was done over all populations, and between pairs of populations. The joint distribution was computed using the `cplot` function in the FDIST2 software package, and  $p$ -values were calculated using the `pv` function. This approach is referred to the FDIST2-like method hereafter.

(iii) Finally, we extended Goldringer & Bataillon’s (2004) approach in order to detect loci exhibiting extreme  $F_{ST}$  values compared to the neutral expectation given the experimental setting and the genetic composition of the initial population. As in the original approach, we simulated the evolution of allele frequencies in three Wright-Fisher populations for 12 generations (which corresponds to the G0-G12 time interval), with each population starting from the same allele frequencies observed in the initial population, and evolving with the same  $N_e$  as the one estimated on the set of presumably neutral markers (see above). Because the available number of individuals from the initial population at G0 was low ( $n = 36$ ), we suspected that the allele frequencies estimated in this population were not very accurate. Conversely, the allele frequencies calculated in the parental population were the exact values because they were assessed on all the 16 parental lines used to obtain the initial population.

Therefore, we generated the distribution of the allele frequencies in G0 by simulating 3000 replicates of a population starting from the allele frequencies of the parental population, and evolving under the  $N_e$  estimated between the parental population and the G0 population ( $N_e = 86$ ). From this distribution of allele frequencies, we simulated the evolution of three diverging Wright-Fisher populations for 12 generations with  $N_e = 150, 136$  and  $40$  for Le Moulon, Toulouse and Vervins, respectively. The Weir & Cockerham's (1984)  $F_{ST}$  estimator  $\theta$  observed at each candidate gene polymorphism was then tested against the simulated distribution of  $\theta$  obtained overall and between pairs of populations from the procedure just described. A *p-value* was assigned to each locus-specific  $\theta$  value, corresponding to the null hypothesis: "population differentiation is the same as expected with evolution under genetic drift with the corresponding  $N_e$  values".

#### *Association studies with flowering time*

Since they were at very low frequencies, heterozygotes were removed from the dataset, in order to avoid under-represented classes of genotypes. The association between the candidate-gene polymorphisms and the variation in flowering time was tested by either considering the effect of each gene polymorphism ( $gene_i$ ) separately:

$$Y_{ijkl} = \mu + gene_i + pop_j + generat(pop)_{jk} + yr_l + \varepsilon_{ijkl} \quad (3a)$$

or by considering, for each gene, the effect of full haplotypes across genomes ( $geneH_i$ ):

$$Y_{ijkl} = \mu + geneH_i + pop_j + generat(pop)_{jk} + yr_l + \varepsilon_{ijkl} \quad (3b)$$

We also tested the effects of the different gene haplotypes simultaneously and their pairwise interactions using the following model:

$$Y_{ijklm} = \mu + geneH1_i + geneH2_j + (geneH1 * geneH2)_{ij} + pop_k + generat(pop)_{kl} + yr_m + \varepsilon_{ijklm} \quad (3c)$$

where  $geneH1_i$  and  $geneH2_j$  are the two gene haplotypes tested simultaneously. Finally, as the gene-by-population interactions were significant in preliminary analyses, we also ran the three ANOVA models in each population separately, discarding the population effect (models 3d, 3e and 3f). The global effect of each gene was also assessed by comparing the determination coefficient ( $R^2$ ) of the above models with the  $R^2$  of the same models without any polymorphism effects.

## Results

### *Evolution of microsatellite markers*

The number of alleles per microsatellite locus ranged from three to 11. We found new alleles in the three populations at G12 that were absent from the parental population. The average diversity in Toulouse and Le Moulon ( $H_e = 0.53$  for the two populations at G12) was at the same level as in the parental population ( $H_e = 0.55$ ). On the contrary, the average diversity in the Vervins population was lower compared to the parental lines ( $H_e = 0.42$ ; Table 1). The overall spatial population differentiation was low but significant in G2 ( $F_{ST} = 0.08$ ,  $CI_{95\%} = [0.05-0.12]$ ,  $p=0.0002$ ) and G7 ( $F_{ST} = 0.08$ ,  $CI_{95\%} = [0.05-0.10]$ ,  $p=0.0002$ ), and increased in G12 ( $F_{ST} = 0.14$ ,  $CI_{95\%} = [0.11-0.17]$ ,  $p=0.0002$ ). Pairwise  $F_{ST}$  estimates showed that Vervins was the most differentiated population (Figure 2). In Vervins and Toulouse populations, the average genotypic disequilibrium measured among pairs of microsatellite loci increased between G2 and G12 (Table 1). In generation 12, 18%, 39% and 32% of the pairwise correlations of gene frequencies were significant ( $p < 0.01$ ) in Le Moulon, Toulouse and Vervins, respectively.

**Table 1 a-** Evolution of diversity and mean pairwise correlations estimated among loci for the 21 microsatellite markers in each population. **b-** Mean genetic diversity estimated from all studied polymorphisms at candidate gene. *S.D.*: Standard deviation; NA: Not available.

		a- Microsatellite markers				b- Candidate gene markers	
Population	Generation	$H_e$	<i>S.D.</i>	$R$	<i>S.D.</i>	$H_e$	<i>S.D.</i>
Parental lines		0.55	0.19	NA	NA	0.33	0.20
Initial population	G0	0.56	0.20	0.19	0.08	0.30	0.19
Vervins	G2	0.45	0.22	0.22	0.12	0.22	0.19
	G7	0.49	0.21	0.21	0.11	0.33	0.22
	G12	0.42	0.21	0.26	0.13	0.30	0.22
Le Moulon	G2	0.52	0.21	0.14	0.05	0.31	0.19
	G7	0.53	0.19	0.16	0.06	0.29	0.20
	G12	0.53	0.22	0.16	0.06	0.27	0.20
Toulouse	G2	0.54	0.20	0.14	0.05	0.31	0.20
	G7	0.54	0.20	0.16	0.06	0.34	0.20
	G12	0.53	0.21	0.21	0.09	0.33	0.20

The effective size for each population was estimated from the temporal variation of allele frequencies between the parental lines and G12. We found  $N_{eg} = 150$  ( $CI_{95\%} = [93-219]$ ) for Le Moulon,  $N_{eg} = 121$  ( $CI_{95\%} = [75-175]$ ) for Toulouse and  $N_{eg} = 40$  ( $CI_{95\%} = [27-54]$ ) for Vervins. Following Goldringer & Bataillon (2004), we simulated three Wright-Fisher populations using the above estimates of effective size, to generate the expected distribution

of the standardized variance of allele frequencies,  $\hat{F}_{c,l}$ . We then tested whether the  $F_{c,l}$  estimated at each microsatellite locus departed from the distribution expected under pure drift. At one locus (*Xgwm261*), the  $F_{c,l}$  estimate was significantly larger than expected under neutrality ( $p = 0.008$ ) in the Toulouse population. We therefore discarded this locus for this population, and obtained a new estimate of effective size in Toulouse:  $N_{eg} = 136$  ( $CI_{95\%} = [82-200]$ ). This estimate of  $N_{eg}$  in Toulouse was further used in subsequent tests of selection for SNPs and haplotypes.

### *Evolution of flowering time*

In all the populations, we found that flowering time was highly heritable. In the population from Le Moulon, the heritability tended to decrease between G2 and G12 (from 0.89 to 0.81), whereas it remained stable in the Vervins and Toulouse populations (0.91 to 0.93 for the two populations). Comparison of means showed that flowering time was not significantly different between the parental lines ( $Means_{Par} = 1222$ ) and the initial population ( $Means_{G0} = 1204$ ) ( $p = 0.07$ ). Considering population and generation effects (ANOVA, model 1), we found a significant population effect and a weaker yet highly significant generation effect within populations (Table 2). In Vervins, flowering time continuously increased ( $p < 0.0001$ ) between G2 ( $LS-Means_{VV-G2} = 1260$ ) and G12 ( $LS-Means_{VV-G12} = 1287$ ). In Le Moulon, flowering time remained stable over generations ( $LS-Means$  value from 1237 in G2 to 1234 in G12). In Toulouse, flowering time decreased significantly ( $p = 0.0005$ ) between G2 ( $LS-Means_{TO-G2} = 1220$ ) and G12 ( $LS-Means_{TO-G12} = 1199$ ).

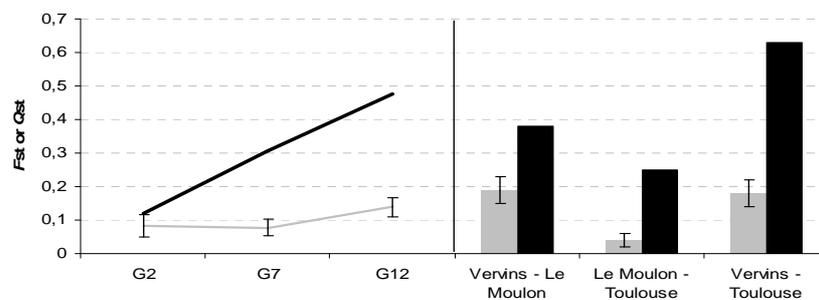
**Table 2:** Results of ANOVA (model (1)) testing the population (Le Moulon, Toulouse, Vervins) and generation (G2, G7, G12) effects on flowering time (HD)

Model	Trait	R <sup>2</sup>	Source of variation	d.f.	Sum of squares	Mean square	F	P
(1)	FT	0.80	pop	2	660805	330402	128.16	<0.0001
			gene (pop)	6	58059	9676	3.75	0.0011
			yr	1	8535813	8535813	3310.86	<0.0001

Population differentiation for flowering time, estimated at each generation, was significant as early as G2 ( $Q_{ST} = 0.10$ ,  $P = 0.0004$ ) and increased until G12 ( $Q_{ST} = 0.48$ ,  $p < 0.0001$ ). Pairwise population comparisons at G12 showed that the Vervins population was the most differentiated from the other two as observed for the microsatellite data. As shown in Figure 2, in G2, population differentiation for flowering time was at the same level as population differentiation inferred from the 21 microsatellite markers. Thereafter, population

differentiation for flowering time increased much more than that differentiation inferred from the microsatellite markers.

**Figure 2:** Population differentiation for the microsatellite markers ( $F_{ST}$ , in grey) and for flowering time ( $Q_{ST}$ , in black). A. Between the three populations at the different generations considered; B. Between pairs of populations at generation 12. 95% bootstrap confidence intervals are shown for  $F_{ST}$  estimates.



### *Evolution of candidate genes*

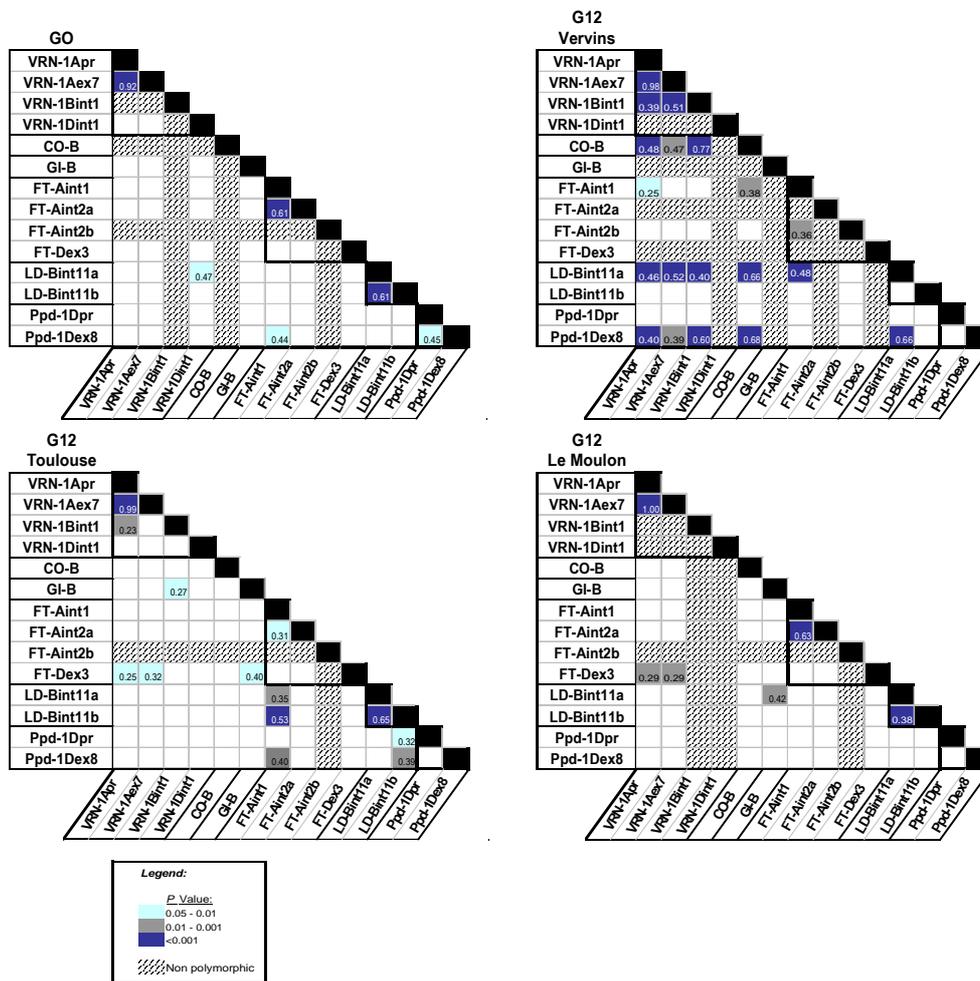
#### *Sequence data for LD, GI and CO*

For *LD*, out of 2000 base pairs sequenced in genomes A, B and D for the 16 parental lines, only two SNPs were identified in genome B. A BLASTN search (National Centre for Biotechnology Information: <http://www.ncbi.nlm.nih.gov/BLAST/>) revealed that the two SNPs were located within a wheat intron, 300 bases from each other, and between two exons that are orthologous to the 11<sup>th</sup> exon and the 12<sup>th</sup> exon of *LD* gene in rice. These two SNPs were further called *LD-Bint11a* and *LD-Bint11b*. For *GI*, out of 1300bp for the 16 parents in the three genomes, a single SNP was found in genome B. A BLASTN search revealed that it was localized within a wheat exon, which is orthologous to the 5<sup>th</sup> exon of the *GI* gene in rice and barley (Dunford et al. 2005). Only 300bp were sequenced for the *CO* gene fragment in the three wheat genomes. A single SNP was observed in genome B. According to the BLASTN search, it was localized within the single intron of the gene, between two exons that are orthologous to the two exons of the *CO* gene in barley, *HvCO3* (Griffiths et al. 2003).

#### *Genetic diversity and linkage disequilibrium*

The average Nei diversity ranged from 0.22 to 0.33 (see Table 1). These values were lower than those estimated from the microsatellite markers, but the two measures were not directly comparable because they were calculated from markers with very different numbers of allele on average. Overall, genetic diversity of candidate genes followed the same trend of evolution as microsatellite markers, except in Le Moulon, in which diversity tended to decrease across generations.

**Figure 3:** Linkage disequilibrium between gene polymorphisms estimated from a measure of correlation between loci. The level of significance of the exact test of gametic disequilibrium is given in the upper side of the matrix, and the value of the correlation is given in the lower side.



Due to the hexaploid nature of the wheat genome, each gene is potentially present in three copies within the genome. Thus, linkage disequilibrium was estimated within each gene copy (i.e. between polymorphisms located in the same gene copy), within each gene (i.e. among polymorphisms located on the different copies of a same gene) and among genes (Figure 3). Linkage disequilibrium within gene copies (estimated for *FT-A*, *VRN-1A*, *LD-B*, and *Ppd-1D*) was high in the initial population and remained significant in all populations after 12 generations of evolution, except for the *Ppd-1D* polymorphisms (in the three populations) and for the *LD-B* polymorphisms in Vervins. Correlations within genes among copies (estimated for *VRN-1* and *FT* genes) became significantly different from zero across generations between *VRN-1A* and *VRN-1B* in Toulouse and Vervins, whereas no linkage disequilibrium was found between the two studied copies of the *FT* gene. Significant correlations among the different genes were also found in the different populations and in particular in the Toulouse and Vervins populations.

*Temporal variation of candidate-genes polymorphisms*

The evolution of *VRN-1* gene polymorphisms in the three populations between the parental population and G12 has been extensively described in Rhoné et al. (2008). The allelic frequencies of the gene polymorphisms within the parental lines and G12 are given in the Online Supplementary Table 1.

**Supplementary table 1:** Allelic frequencies estimated in the parental population and in the three populations in G12.

<i>Gene polymorphism</i>	<i>Allele</i>	<i>Parental</i>			
		<i>lines</i>	<i>Vervins</i>	<i>Le Moulon</i>	<i>Toulouse</i>
<i>VRN-1Apr</i>	1	0.06	0.00	0.00	0.05
	2	0.69	0.32	0.88	0.65
	3	0.00	0.12	0.00	0.00
	4	0.00	0.57	0.00	0.00
	5	0.25	0.00	0.13	0.30
<i>VRN-1Aex7</i>	1	0.69	0.42	0.88	0.65
	2	0.31	0.58	0.13	0.35
<i>VRN-1Bint1</i>	1	0.94	0.51	1.00	0.65
	2	0.06	0.49	0.00	0.35
<i>VRN-1Dint1</i>	1	1.00	1.00	1.00	0.97
	2	0.00	0.00	0.00	0.03
<i>Ppd-1Dpr</i>	1	0.44	0.83	0.63	0.79
	2	0.56	0.18	0.38	0.21
<i>Ppd-1Dex8</i>	1	0.63	0.74	0.38	0.39
	2	0.38	0.26	0.63	0.61
<i>FT-Aint1</i>	2	1.00	0.66	1.00	1.00
	4	0.00	0.34	0.00	0.00
<i>FT-Aint2a</i>	8	0.50	0.21	0.58	0.43
	9	0.50	0.79	0.42	0.57
<i>FT-Aint2b</i>	3	0.69	1.00	0.66	0.93
	4	0.31	0.00	0.34	0.07
<i>FT-Dex3</i>	3	0.81	1.00	0.91	0.95
	4	0.19	0.00	0.09	0.05
<i>LD-Bint11a</i>	2	0.13	0.65	0.36	0.38
	4	0.88	0.35	0.64	0.62
<i>LD-Bint11b</i>	1	0.69	0.97	0.80	0.60
	2	0.31	0.03	0.20	0.40
<i>CO-B</i>	2	0.13	0.54	0.02	0.02
	4	0.88	0.46	0.98	0.98
<i>GI-B</i>	2	0.88	1.00	0.60	0.88
	4	0.13	0.00	0.40	0.12

**Supplementary table 2:** Allelic combination of the multilocus frequencies estimated in the parental population and in the three populations in G12.

<i>haplotype number</i>	<i>VRN-1Apr</i>	<i>VRN-1Aex8</i>	<i>VRN-1Bint1</i>	<i>VRN-1Dint1</i>	<i>Parental population</i>	<i>Vervins</i>	<i>Le Moulon</i>	<i>Toulouse</i>
h1	1	2	2	1	0.06	0.00	0.00	0.05
h3	2	1	1	1	0.69	0.29	0.88	0.37
h4	2	1	1	2	0.00	0.00	0.00	0.03
h5	2	1	2	1	0.00	0.02	0.00	0.25
h7	2	2	2	1	0.00	0.01	0.00	0.00
h8	5	2	1	1	0.25	0.00	0.13	0.25
h10	5	2	2	1	0.00	0.00	0.00	0.05
h11	3	1	1	1	0.00	0.05	0.00	0.00
h12	3	1	2	1	0.00	0.07	0.00	0.00
h14	4	2	1	1	0.00	0.17	0.00	0.00
h15	4	2	2	1	0.00	0.40	0.00	0.00

<i>FT-Aint1</i>	<i>FT-Aint2a</i>	<i>FT-Aint2b</i>	<i>FT-Dex3</i>	<i>Parental population</i>	<i>Vervins</i>	<i>Le Moulon</i>	<i>Toulouse</i>
8	3	2	3	0.19	0.34	0.24	0.24
8	3	2	4	0.00	0.01	0.00	0.00
8	4	2	3	0.19	0.00	0.18	0.33
8	4	2	4	0.13	0.00	0.06	0.02
9	3	2	3	0.44	0.59	0.50	0.42
9	3	2	4	0.06	0.01	0.03	0.00
9	3	4	3	0.00	0.05	0.00	0.00

<i>LD-Bint11a</i>	<i>LD-Bint11b</i>	<i>Parental population</i>	<i>Vervins</i>	<i>Le Moulon</i>	<i>Toulouse</i>
2	1	0.13	0.65	0.36	0.38
4	1	0.56	0.32	0.43	0.22
4	2	0.31	0.03	0.20	0.40

<i>Ppd-1Dpr</i>	<i>Ppd-1Dex8</i>	<i>Parental population</i>	<i>Vervins</i>	<i>Le Moulon</i>	<i>Toulouse</i>
1	1	0.44	0.23	0.36	0.63
1	2	0.19	0.14	0.05	0.11
2	1	0.00	0.39	0.43	0.21
2	2	0.38	0.23	0.16	0.05

Using Goldinger & Bataillon (2004), we found two, four and six gene polymorphisms in Le Moulon, Toulouse and Vervins respectively, that departed significantly from the expected

neutral distribution between the parental lines and G12 (Table 3a). Only one SNP (LD-Bint11a) was an outlier in all three populations, and two polymorphisms (VRN-1B and FT-Aint2b) were outliers in two populations (Vervins and Toulouse). Overall, for each candidate gene considered here, at least one outlier was detected in one population.

#### *Divergent selection of genes polymorphisms*

In order to discard microsatellite markers behaving as outliers loci, we ran both the DETSEL and FDIST2-like methods on the microsatellite dataset, with a strict stepwise mutation model ( $\mu = 0.005$ ). The effective population size of the ancestral population (before divergence) was set to  $N_e = 100$ . We fixed the maximum allele frequency (the frequency of the most frequent allele) in the simulations to 0.99. One locus, *Xgwm664*, was identified as an outlier in the pairwise comparison Le Moulon – Vervins ( $p < 2 \cdot 10^{-5}$ ). Both *Xgwm664* and *Xgwm480* had extremely low heterozygosity (0.058 and 0.037, respectively). These two loci were therefore discarded in subsequent analyses. The relative bias for global  $F_{ST}$  values (the relative difference between the observed values and the average over  $5 \cdot 10^5$  simulations) was less than 0.5%.

We then ran the DETSEL, the FDIST2-like and the modified Goldringer and Bataillon's methods on the candidate-gene polymorphisms. For both the DETSEL and FDIST2-like methods, we used a  $K$  allele mutation model with  $\mu = 10^{-5}$  for SNP polymorphisms, and  $\mu = 0.01$  for haplotypes, as recombination events were considered as equivalent to mutations. As shown in Table 3b, we detected more polymorphisms as potentially selected with the modified Goldringer & Bataillon's method, as compared to the DETSEL and FDIST2-like methods. All polymorphisms found to be under selection with the DETSEL and the FDIST2-like methods were also pointed out by the modified Goldringer & Bataillon's method except *VRN-1A pr*, *CO-B* and *GI-B*, detected as outlier in the Le Moulon-Toulouse comparison with the DETSEL method only. Considering the number of detected outliers overall, Le Moulon and Vervins appeared as the most divergent populations. As shown by both the DETSEL and the modified Goldringer and Bataillon's methods, only one gene haplotype (*VRN-1ABD*) departed from neutrality in the comparison between Le Moulon and the two other populations.

#### *Temporal vs. spatial selection comparison*

As can be seen by comparing Tables 3a and 3b, the outlier loci detected by either spatial method in G12 were also found as outliers at least in one of the populations when using the Goldringer & Bataillon's (2004) temporal approach, except for *VRN-1A<sub>ex7</sub>*: for instance *CO-B*

was presumably selected in Vervins (stronger temporal variation of allele frequencies than expected), and this locus was also more differentiated than expected when comparing Vervins and the other two populations (at least considering the modified Goldringer & Bataillon's method). In contrast, some loci detected as outliers with the Goldringer & Bataillon's (2004) temporal approach were not found to be under spatial divergent selection (see, for example, *LD-B<sub>int11a</sub>* or *Ppd-1D<sub>pr</sub>*). This may be due to the fact that for these two polymorphisms, the same alleles were seemingly selected over generations in all the populations (uniform selection; see the Online Supplementary Tables 1 and 2). Similarly, at the *VRN-1B<sub>int1</sub>* locus, the same allele was presumably selected in the Vervins and Toulouse populations, and the genetic differentiation between each of these populations and Le Moulon was therefore exacerbated at this locus.

**Table 3:** Significant temporal and spatial selection of the gene polymorphisms. **3a** Results from the tests of the null hypothesis that the change in allele frequency that occurred between the parental population and the populations sampled at generation G12 was solely due to drift. **G12. 3b:** Results from the tests of the null hypothesis that the genetic differentiation among populations at generation G12 is solely due to drift. The three approaches considered refer to: (1) Vitalis et al.'s (2001) DETSEL method, extended for the simultaneous divergence of three populations; (2) the FDIST2-like method (an analysis of the DETSEL divergence model using the joint distribution of  $F_{ST}$  and heterozygosity); (3) an extension of Goldringer & Bataillon's (2004) approach using a model that simulates the divergence of three populations evolving from a known genetic pool (see the main text for details). Only significant  $p$ -values are reported. Full haplotypes are indicated in light grey.

gene polymorphism	Chromosomic location	a- Temporal selection (Parents-G12)			b- Spatial selection (G12)								
		Vervins	Le Moulon	Toulouse	Vervins- Le Moulon			Le Moulon- Toulouse			All Populations		
					(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
<i>VRN-1A<sub>pr</sub></i>	5A	☆			★	★	★	★				★	★
<i>VRN-1A<sub>ex7</sub></i>	5A						★						
<i>VRN-1B<sub>int1</sub></i>	5B	☆☆		☆☆			★★	★	★★				★
<i>VRN-1D<sub>int1</sub></i>	5D	NP	NP			NP							
<i>VRN-1ABD</i>		☆		☆	★★	★	★★	★	★				★
<i>Ppd-1D<sub>pr</sub></i>	2D			☆☆									
<i>Ppd-1D<sub>ex8</sub></i>	2D												
<i>Ppd-1D</i>		☆	☆☆	☆☆									
<i>FT-A<sub>int1</sub></i>	7A	☆☆☆	NP	NP		★★		NP			★★	★	★★
<i>FT-A<sub>int2a</sub></i>	7A												
<i>FT-A<sub>int2b</sub></i>	7A	☆		☆									
<i>FT-D<sub>ex3</sub></i>	7D												
<i>FT-AD</i>													
<i>LD-B<sub>int11a</sub></i>	3B	☆☆	☆	☆									
<i>LD-B<sub>int11b</sub></i>	3B												
<i>LD-B</i>		☆		☆									
<i>CO-B</i>	5B	☆				★★	★				★	★	★
<i>GI-B</i>	3B		☆☆			★	★						

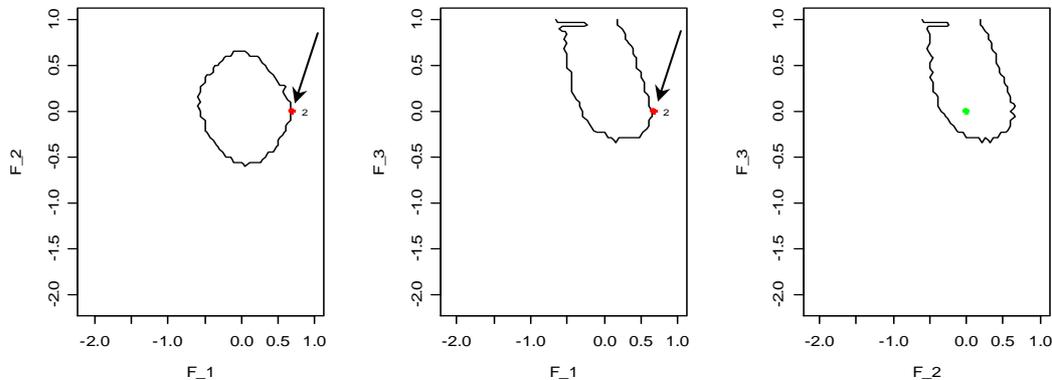
NP: Non polymorphic in the two populations compared

☆  $P$ -value <0.05

☆☆  $P$ -value <0.01

☆☆☆  $P$ -value <0.001

Supplementary figure: Examples of analyses performed with DETSEL method extended for the simultaneous divergence of three populations: detection of divergent selection for the *VRN-1ABD* haplotype. The envelope of the simulated values of F1 and F2 are plotted. The simulation were performed for  $K = 10$  allele mutation model with  $\mu = 0.01$  for haplotypes. The arrows point the position for *VRN-1ABD* when it is significant.



#### Gene polymorphisms or haplotypes association with flowering time

**Table 4:** Contribution of each studied gene polymorphisms to flowering time variation using either the three population together (ANOVA model 3a for SNP data and 3b for haplotype data) or each population separately (model 3d and 3e). Number indicated the determination coefficient ( $R^2$ ) of the model when the effect of the tested gene is significant. *NP*: Non polymorphic in the population considered.

		$R^2$ ANOVA without gene effect	Model 3a (SNPs) or 3d (haplotypes) 0,794	Model 3b (SNPs) or 3e (haplotypes) 0,749    0,840    0,770		
Chromosomic location	gene polymorphism		3 populations	Vervins	Le Moulon	Toulouse
Chr 5	5A	<i>VRN-1A pr</i>	0,806	0,762		0,784
	5A	<i>VRN-1A ex7</i>	0,805	0,763		0,781
	5B	<i>VRN-1B int1</i>		0,758		0,775
	5D	<i>VRN-1D int1</i>	0,803	NP	NP	0,791
		<i>VRN-1ABD</i> (haplotype)	0,824	0,801		0,808
Chr 2	2D	<i>Ppd-1D pr</i>				
	2D	<i>Ppd-1D ex8</i>		0,748		0,787
		<i>Ppd-1D</i> (haplotype)				
Chr 7	7A	<i>FT-A int1</i>	0,800	0,753	NP	NP
	7A	<i>FT-A int2a</i>		0,773	0,852	0,775
	7A	<i>FT-A int2b</i>	0,800	NP		0,781
	7D	<i>FT-D ex3</i>	0,800	0,755	0,845	
		<i>FT-AD</i> (haplotype)	0,807	0,778	0,862	0,787
Chr 3	3B	<i>LD-B int11a</i>	0,801	0,766		
	3B	<i>LD-B int11b</i>				0,773
		<i>LD-B</i> (haplotype)		0,763		0,774
Chr 5	5B	<i>CO-B</i>	0,803	0,764		
Chr 3	3B	<i>GI-B</i>		NP	0,841	0,775



The genetic associations between candidate gene polymorphisms and flowering time are summarized in Table 4. Except for *Ppd-ID<sub>pr</sub>*, all the candidate genes studied were significantly associated with variation in flowering time. However, the strength of the association depended on the polymorphism and the population considered. *VRN-1* and *FT* were found to be far more strongly associated with the trait than *Ppd-1*, particularly when full haplotypes (across genomes) were considered, suggesting that epistatic interactions among gene copies across genomes may play a significant role. Among the newly described polymorphisms in *LD*, *GI* and *CO*, the *CO-B* polymorphism showed the more important effect on the variation in flowering time. Interactions among genes were also found to contribute significantly to the trait variation (Table 5). Comparing  $R^2$  values between full (ANOVA model 3b) and partial models (ANOVA model 3b without gene effects) showed that overall, 3% of the total phenotypic variation was explained by *VRN-IABD* variations alone (Table 4). Nevertheless, 4% of the total phenotypic variation was explained by either the *VRN-IABD*  $\times$  *Ppd-ID* or the *VRN-IABD*  $\times$  *LD-B* interaction, and 5% of the total variation was explained by the *VRN-IABD*  $\times$  *FT-AD* interaction (Table 5). Within populations, including *VRN-IABD*  $\times$  *FT-AD* interaction into the ANOVA explained as much as 9% of the total phenotypic variation in Vervins and Toulouse (Table 5), compared to 4-5% of the total variation without this interaction (Table 4). In contrast to the other populations, the total phenotypic variation in Le Moulon was mostly explained by genetic variation at the *FT* gene (2% of total phenotypic variation), while the effect of *VRN-1* was not significant. At best, 3% of the total phenotypic variation observed in this population was explained by the ANOVA which included the *VRN-IABD*  $\times$  *FT-AD* interaction.

**Supplementary table 3:** Contribution of haplotype variation and their interaction to flowering time using either the three population together (ANOVA model 3c) or each population separately (model 3f). Number indicated the determination coefficient ( $R^2$ ) of the model when the effect of the tested gene is significant. *NP*: Non polymorphic in the population considered; *NR*: Interaction non realizable.

<i>R</i> <sup>2</sup> without gene effect	Model 3c		Model 3f		P-value
	3 populations	Vervins	Le Moulon	Toulouse	
<i>VRN-1ABD</i>	0,824	0,801		0,808	<0.0001
<i>Ppd-1D</i>					
<i>FT-AD</i>	0,807	0,778	0,862	0,787	<0.0001
<i>LD-B</i>		0,763		0,774	<0.01
<i>CO-B</i>	0,803	0,764			<0.001
<i>GI-B</i>		NP	0,841	0,775	<0.05
<i>VRN-1ABD</i>	0,832	0,823	0,840	0,826	<0.0001
<i>Ppd-1D</i>					
Interaction					<0.05
<i>VRN-1ABD</i>	0,846	0,837	0,871	0,856	<0.0001
<i>FT-AD</i>					
Interaction					<0.0001
<i>VRN-1ABD</i>	0,834	0,819	0,847	0,846	<0.0001
<i>LD-B</i>					
Interaction					<0.0001
<i>VRN-1ABD</i>	0,829	0,813	0,840	0,813	<0.0001
<i>CO-B</i>					
Interaction			NR		
<i>VRN-1ABD</i>	0,825	0,801	0,843	0,813	<0.0001
<i>GI-B</i>		NP			
Interaction		NR			
<i>Ppd-1D</i>	0,809	0,781	0,868	0,803	<0.0001
<i>FT-AD</i>					
Interaction					<0.0001
<i>Ppd-1D</i>	0,799	0,776	0,838	0,784	<0.0001
<i>LD-B</i>					
Interaction					<0.05
<i>Ppd-1D</i>	0,805	0,773	0,835	0,775	<0.0001
<i>CO-B</i>					
Interaction					<0.0001
<i>Ppd-1D</i>	0,796	0,754	0,837	0,784	<0.0001
<i>GI-B</i>		NP			
Interaction		NR			
<i>FT-AD</i>	0,810	0,787	0,867	0,804	<0.0001
<i>LD-B</i>					
Interaction					<0.0001
<i>FT-AD</i>	0,812	0,787	0,861	0,787	<0.0001
<i>CO-B</i>					
Interaction			NR		
<i>FT-AD</i>	0,808	0,778	0,863	0,789	<0.0001
<i>GI-B</i>		NP			
Interaction		NR			
<i>LD-B</i>	0,804	0,769	0,839	0,776	<0.0001
<i>CO-B</i>					
Interaction					<0.0001
<i>LD-B</i>	0,801	0,763	0,840	0,779	<0.0001
<i>GI-B</i>		NP			
Interaction		NR			
<i>CO-B</i>	0,803	0,764	0,840	0,777	<0.0001
<i>GI-B</i>		NP			
Interaction		NR	NR		

## Discussion

### *Neutral evolution of the wheat populations*

As previously found in the wheat DM populations (Goldringer et al. 2001; Rhoné et al. 2007; Raquin et al. 2008), effective population sizes estimated from neutral markers were found to be much lower than the demographic population sizes estimated on the basis of the minimal census number of seeds sown every year to constitute the next generation ( $N \approx 2650$ ). In a meta-analysis including 192 experiments on natural populations, Frankham (1995) showed that the ratio  $N_e/N$  was about 0.10. Siol et al. (2007) found a ratio between 0.05 and 0.10 in the highly selfing species, *Medicago truncatula*, explaining this result by a variation in the contribution of the different lines to the following generations. The  $N_e/N$  ratio observed in Le Moulon and Toulouse were of the same magnitude (0.06 and 0.05 respectively). As the demographic effective sizes of the populations were controlled over generations, the low  $N_e$  values cannot be attributed to demographic fluctuations, but rather may be the signature of variation in the reproductive contribution (Raquin et al. 2008).

The Vervins population showed the lowest  $N_e$  value (40) and thus the lowest  $N_e/N$  ratio (0.01). Among the different time intervals available to estimate  $N_e$ , the G0-G2 interval gave the lowest effective size in this population ( $N_e = 9$ ). This observation, combined with the reduction of diversity observed at the microsatellite markers between G0 ( $H_e = 0.56$ ) and G2 ( $H_e = 0.45$ ) and the highest global correlation found between the unlinked microsatellite markers in G2 ( $R = 0.22$  in Vervins,  $R = 0.14$  in Toulouse and Le Moulon), suggested that the population might have experienced a severe bottleneck at its foundation. This scenario had already been suggested by Paillard et al. (2000), who followed the evolution of resistance genes to a wheat pathogen, the powdery mildew, between the parents and G10 in Vervins. They observed strong stochastic changes in allele frequencies leading eventually to the loss of some known resistance alleles, which was at odd with the fact that this population was submitted to strong pathogenic pressures.

### *Flowering time differentiation as a response to local adaptation*

Populations grown in contrasted environments showed increasing differentiation for flowering time over the studied generations. A latitudinal gradient of flowering time was observed, the populations from the North flowering later than the population from the South. Flowering time is known to be an important adaptive trait ensuring that reproduction occurs under

favourable environmental conditions. In the studied populations, divergent evolution of flowering time could be linked to local climatic conditions: in Southern France, plants with a late flowering date are exposed to heat and drought stresses in early summer. In contrast, in the North where the winter season is long and cold, genotypes that wait for the warmer season to initiate flowering may be favoured, whereas genotypes that flower too early in spring may be exposed to frost injury. Divergent evolution of flowering time along environmental gradients in relation to local adaptation has been observed in several plant species such as *Beta vulgaris ssp maritima* (Boudry et al. 2002), *Arabidopsis thaliana* (Stinchcombe et al. 2004), or *Arabidopsis lyrata* (Sandring et al. 2007). Early flowering appears to be selected for in environments with a short or unpredictable growing season (Roux et al. 2006). In this study, we found that  $Q_{ST}$  estimated for flowering time over generations in a common garden experiment increased much more rapidly than  $F_{ST}$  estimated on 21 neutral microsatellite markers. In G7 and G12, the  $Q_{ST}$  value was higher than the  $F_{ST}$  value, which is a signature of divergent selection acting on this trait to achieve local adaptation. Indeed, under neutrality, for a quantitative trait with purely additive genetic basis,  $Q_{ST}$  is supposed to be equal to  $F_{ST}$  (Spitze 1993; Whitlock 1999). Based on this assumption,  $Q_{ST}$  greater than  $F_{ST}$  is interpreted as the consequence of diversifying selection acting on the trait, favouring different phenotypes in different environments; in contrast,  $Q_{ST}$  smaller than  $F_{ST}$  is interpreted as a consequence of stabilizing selection where the same phenotypes are selected for in different environments. A comparison of  $F_{ST}$  with  $Q_{ST}$  is widely used to detect local adaptation of populations evolving in contrasted environments. However, the use of this comparison to infer selection has been debated recently. In particular, biases altering estimations of marker differentiation were suspected to invalidate the detection of selection on quantitative traits. Merilä & Crnokrak (2001) challenged the assumption that molecular markers chosen to estimate  $F_{ST}$  behave neutrally: non-coding markers can indeed potentially be affected by indirect selection acting on a neighbouring gene through the hitch-hiking effect (for example Maynard Smith & Haigh 1974), leading to inflated  $F_{ST}$ 's. Conversely,  $F_{ST}$  could be under-estimated when measured using highly polymorphic markers such as microsatellites. The high intra-population diversity associated with their mode of mutation presents a high risk of homoplasy and might result in an under-estimation of the among-population variance (Flint et al. 1999). In our study, markers used to estimate the neutral  $F_{ST}$  have first been tested for spatially neutral evolution with a method of outlier detection and those presenting deviation from neutral expectation were discarded from the analyses. Moreover, the  $F_{ST}$  for highly polymorphic microsatellite loci does not seem to be under-estimated since the value observed on less variable RFLP

markers after ten generations (G10) of evolution between the Le Moulon and Toulouse populations was quite close ( $F_{ST} = 0.03$  in Enjalbert et al. 1999) to the  $F_{ST}$  estimated in G7 ( $F_{ST} = 0.02$ ) or in G12 ( $F_{ST} = 0.04$ ) in our experiment.

In addition, non-additive effects (i.e. epistatic and dominance effects) and inbreeding have been suspected to affect  $Q_{ST}$  estimation (Lynch et al. 1999; Whitlock 1999; Merilä & Crnokrak 2001; Lopez-Fanjul et al. 2003; Lopez-Fanjul et al. 2007). Yet both theoretical (Le Corre & Kremer 2003; Goudet & Büchi 2006; Goudet & Martin 2007) and an experimental (Porcher et al. 2006) studies showed that  $Q_{ST} > F_{ST}$  (as it is the case in the present study) is a robust evidence of divergent selection (Leinonen et al. 2008). Finally, using both simulations and empirical data, O'Hara & Merilä (2005) showed that analyzing less than 20 populations can affect the precision of  $Q_{ST}$  estimation due to a poor estimation of the among-population variance. However, as suggested by Evanno et al. (2006), this problem is limited when the genetic variance is evenly distributed among the different populations, as it is the case in our study.

#### *Genetic architecture of flowering time adaptive variation*

Among the candidate-gene polymorphisms, some showed a clear association with flowering time (*VRN-1* and *FT* genes) especially under their haplotypic form, whereas others showed more subtle associations through their interaction with other genes (*LD* and *Ppd* interacting with *VRN-1*) or through their association with variation in flowering time in only one population (*CO* in Vervins). As a result, the detection of diversifying selection clearly depends on the effect of the gene on the selected character and also on its initial diversity and its interaction with other loci. In our study, we used both temporal and spatial methods of detection of selection. Except for one locus, these methods provided complementary information, demonstrating the high power of detection of these methods. Moreover, the effect of each studied gene has been estimated on the selected quantitative trait itself, considering the effects of the interactions between the different copies of each gene (specific to the allohexaploid nature of the wheat genome) and also those existing among genes.

Among and within populations, two genes out of six (*VRN-1* and *CO*) were strongly associated with flowering time variation, and simultaneously detected as being temporally and spatially selected. *VRN-1* had the strongest impact on the trait variation, especially under its haplotypic form (Table 4). Considering each *VRN-1* polymorphism separately, out of the four loci used in this study, only *VRN-1A<sub>pr</sub>* was simultaneously related to earliness and spatio-

temporally selected. Detection of temporal selection in the Vervins population was due to the increase in frequency of two new alleles not previously present within the parental lines but appeared specifically in the Vervins population by mutation or accidental migration in G2. Interaction effects between copies of *VRN-1* have already been reported to determine both winter growth habit (genotypes which need a period of exposure to low temperature called vernalization to initiate flowering) and spring growth habit (insensitivity to vernalization) in those populations (Rhoné et al. 2008) and more generally in the hexaploid wheat (Stelmakh 1987). A spring allele at one of the three *VRN-1* copies leads to a spring growth habit genotype. Our experimental work illustrates that epistatic interaction between genes may hide the response to divergent selection at individual locus. Indeed, we found that *VRN-1B<sub>int1</sub>* was selected evenly in both the Vervins and Toulouse populations, with an increased frequency of the *Vrn1B<sub>int1</sub>-2* allele in the two populations despite their contrasted climatic conditions. However this allele was selected in association with different alleles at the *VRN-1A* gene leading to late flowering in the population of Vervins and early flowering in the population of Toulouse. Thus, the *VRN-1B<sub>int1</sub>* locus was not detected to be under divergent selection in the comparison, even though it seemed to play a key role in the response to divergent selection of flowering time in those two populations.

The previous observations and the level of linkage disequilibrium that increased across generation between the copies of the *VRN-1* gene located on different chromosomes suggested that selection acted on multilocus allelic combinations in relation to the epistatic interactions found between the copies of this gene. To test this hypothesis, selection was tested on the multilocus haplotype *VRN-1ABD* combining the different alleles observed at the three copies of the gene. In the Le Moulon population, the haplotypic form of the gene was not found to be temporally selected and was not associated with flowering time. This was probably due to its low diversity in that population. Indeed, only two *VRN-1ABD* haplotypes were found in G12 and those two haplotypes were associated with the same value for flowering time ( $h_3 = 1245$ ;  $h_5 = 1247$ ). In relation with its high effect in Toulouse and Vervins populations, *VRN-1ABD* was found to be clearly temporally and spatially selected when compared to the initial population and to Le Moulon respectively, but significant divergence was not detected between Toulouse and Vervins, whichever method was used. However, numerous and different allelic combinations were promoted in these two environments (Table 2 supplementary material, Rhoné et al. 2008), leading to a rather low value of genetic differentiation in G12 ( $F_{ST} = 0.168$ ), only slightly higher than the value obtained on the microsatellite markers ( $F_{ST} = 0.139$ ,  $CI_{95\%} = [0.110-0.167]$ ). This result might

be explained by the high intra-population variance that tends to increase the overall variance compared with the among-population variance, leading to a low value of differentiation. This observation has been discussed elsewhere (Charlesworth 1998; Balloux et al. 2000) and could present a serious disadvantage of using  $F_{ST}$  measure to detect selection in such cases.

In agreement with its effect on flowering time variation, *CO-B* was found to be selected both over generations and divergently among populations (methods 1 and 3) in the Vervins population. In contrast to the Toulouse and Le Moulon populations, the allele conferring late flowering increased in frequency in this population. One may suspect that the high linkage disequilibrium value observed between *CO-B* and *VRN-1B* in the Vervins population ( $R^2 = 0.77$ , Figure 3) is probably the best explanation of the signal detected for this locus in that population..

The *FT* gene has recently been shown to be involved in the vernalization pathway of flowering time in wheat (Yan et al. 2006; Bonnin et al. 2008). In Le Moulon population, it was the only studied gene clearly associated with flowering time, especially in its haplotypic form, but *FT-AD* variation explained 2% of the total phenotypic variation only (instead of 4-5% for *VRN1-ABD* in Toulouse and Vervins respectively in Table 4). As suggested by Hall et al. (2007), this low-to-moderate phenotypic effect may explain the absence of evidence for selection using haplotypic data. Moreover, the *FT* gene is involved in the trait architecture mainly through its interaction with *VRN-1* gene in the Toulouse and Vervins populations and as shown before, this situation could hide the response to selection. However, one polymorphic site from the genome A, *FT-A<sub>int1</sub>*, showed a spatio-temporal signal whereas its phenotypic effect was negligible (0.4% in Vervins in Table 4) This polymorphism was not found in the parental population, but appeared in Vervins by mutation or by accidental migration in G2. This result is in accordance with Teshima et al. (2006), who showed that a favorable allele appearing in a population by mutation or migration is more easily recognized as selected than a favorable allele coming from the standing variation.

The three genes *PpD*, *GI* and *LD* were only slightly associated with flowering time, and except for *GI* in population comparisons involving Le Moulon, no spatial selection was detected, whereas temporal selection was apparent. Given that *PpD* is a major gene governing flowering time through its implication in the photoperiod pathway (Worland et al. 1998; Beales et al. 2007), this result suggests that the populations were not differentiated for their photoperiod sensitivity, probably due to the small difference of day length between the two

more extreme sites studied (Goldringer et al. 2006). However this conclusion needs further investigation since we only studied the evolution of polymorphisms identified in one copy of the gene (*Ppd-D*). Even if the D copy is known to have the strongest effect on photoperiod sensitivity (Cockram et al 2007), variations at the A or B copies of the gene could also contribute to the variation of flowering time found in our populations. *GI* and *LD* polymorphisms were not previously described and further association and expression studies should be conducted to ascertain their role in the wheat flowering pathway.

#### *No evidence of selection, or no selection?*

There are several potential explanations of the absence of divergent selection detected at several genes which were thought to be involved in variation of flowering time. First, despite the use of three different approaches to detect selection, selection acting on the genes could have been missed. In a recent study on phenology in European aspen, Hall et al. (2007) did not detect genetic differentiation at putative QTLs, i.e. SNPs in which they could observe clinal variation. They concluded that their results were in accordance with theoretical predictions, but also questioned the power of “ $F_{ST}$  outlier” methods to detect QTLs with low-to-moderate phenotypic effects, citing the scarcity of empirical studies available to investigate this point. Previous theoretical work based on analytical (Latta 1998) or simulations studies (Le Corre & Kremer 2003) have shown that heterogeneous selection on a trait governed by many QTL with pure additive effects could generate positive covariances between QTL (or linkage disequilibrium) leading to a QTL-based  $F_{ST}$  better described by  $F_{ST}$  at neutral loci rather than by  $Q_{ST}$  at the quantitative traits themselves. In the present study, high linkage disequilibrium were found in particular in the population of Toulouse and Vervins between the *LD-B*, *FT-A* and *Ppd-D* genes (Figure 2) which could explain why no spatial selection was detected at these genes. Linkage disequilibrium found among loci are usually associated to epistatic interaction found between these genes and the *VRN-1* gene, suggesting that selection acted on multilocus allelic combinations and such a mechanism could hide the response to selection at individual loci.

A second explanation is that the studied genes were not selected at all in the wheat populations. Roux et al. (2006) suggested that many genes involved in flowering time showed pleiotropic effects. They argued that the number of genes affected by selection for this trait may be limited by such pleiotropic effects. For example, Börner et al (1993) suspected that the *Ppd* gene could have a pleiotropic effect by acting both on flowering time insensitivity and

plant height reduction. Previous studies characterizing the evolution of the wheat experimental populations (Raquin et al. 2008) have shown that mean plant height increased in all the populations in response to competition for light. This observation could explain the temporal selection observed at this gene leading to the promotion of the same alleles in the three populations and the absence of divergent selection detected despite its implication in variation in flowering time.

## Conclusions

In the present study, we combined an indirect approach, identifying a quantitative trait that is a likely target of selection based on its genetic divergence ( $Q_{ST}$ ), with a direct test for divergent selection at the underlying QTLs, which were tested for their association with the trait variation. We showed that detecting divergent selection of genes underlying phenotypic variation among populations developing local adaptation is not a trivial exercise, especially for a polygenic character. High linkage disequilibrium in a bottlenecked population could lead to the detection of false-positive genes, especially in a highly autogamous species such as wheat. Our experimental results also illustrated the fact that many genetic changes contributing to adaptation may be missed because the signature of divergent selection is not recognized (Le Corre & Kremer 2003; Teshima et al. 2006). Among the three methods used in this study, method (3) (derived from Goldringer & Bataillon) was the most efficient to detect selection. This is because this method takes most of the information from the studied populations into account (namely, the allele frequencies in the initial population). The DETSEL and FDIST2-like methods (which are more widely used) appeared to be highly conservative and the only genes detected as selected are the major genes that have a high impact on the trait. With low-to-moderate phenotypic effect, or when covariance among genes occurs, even combining several spatial methods of outlier loci detection appeared inefficient. In such a case, it seems necessary to also test for epistatic interactions between genes and for selection of allelic combination at several loci. Test for homogeneous selection among population using temporal methods appeared also as an efficient method. However such approaches imply that both the history of the populations and the target of selection are known, which is unlikely in natural populations. In such populations, applying a genome scan with “ $F_{ST}$  outlier” methods is still a good prerequisite to detect and isolate major QTLs involved in the adaptive population differentiation.

## **Acknowledgements:**

The authors would like to thank, N. Galic, D. Madur, S. Mezi, C. Remoué and C. Ridel for their technical assistance; S. Praud and the Biogemma society for the genotyping of Ppd-1D gene, G. Boutet and the INRA genotyping platform of Clermont-Ferrand for the microsatellite genotyping of the populations and J. Dawson for careful reading of the manuscript. Part of this work was carried out by using the resources of the Computational Biology Service Unit from the Museum National d'Histoire Naturelle (MNHN) which was partially funded by Saint Gobain. This work was financially supported by the French Bureau des Ressources Génétiques (BRG) and by the Institut National de la Recherche Agronomique (INRA). This work is a part of B. Rhoné's PhD dissertation, supported by the Conseil Régional d'Ile-de-France.

## LITERATURE CITED

- Balloux, F., H. Brüner, N. Lugon-Moulin, J. Hausser, and J. Goudet. 2000. Microsatellites can be misleading: an empirical and simulation study. *Evol.* 54(4):1414-1422.
- Beales, J., A. Turner, S. Griffiths, J.W. Snape, and D.A. Laurie. 2007. A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Gen.* 115:1432-2242.
- Beaumont, M.A., and R.A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B.* 263:1619-1626.
- Beaumont, M.A., and D.J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13:969-980.
- Beaumont, M.A. 2005. Adaptation and speciation: what can  $F_{ST}$  tell us? *Tr. Ecol. Evol.* 20:35-440.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France), Available at <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>.
- Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol. Biol. Evol.* 23:773-783.
- Bonin, A., D. Ehrich, and S. Manel. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol. Ecol.* 16:3737-3758.
- Bonnin, I., J.M. Prosperi, and I. Olivieri. 1996. Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): a comparative analysis of population structure. *Genet.* 143:1795-1805.
- Bonnin, I., M. Rousset, D. Madur, P. Sourdille, C. Dupuits, D. Brunel, and I. Goldringer. 2008. FT genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. *Theor. Appl. Gen.* 116:383-394.
- Börner, A., A.J. Worland, J. Plashke, E. Shuman, and C.N. Law. 1993. Pleiotropic effects of genes for reduced height (Rht) and day-length insensitivity (Ppd) on yield and its components for wheat growth in Middle Europe. *Plant Breed.* 11:204-216.
- Boudry, P., H. McCombie, and H. Van Dijk. 2002. Vernalization requirement of wild beet *Beta vulgaris* ssp. *maritima*: among population variation and its adaptive significance. *J. Ecol.* 90:693-703.
- Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* 15(5):538-543.
- Cockram, J., H. Jones, F.J. Leigh, D. O'Sullivan, W. Powell, D.A. Laurie, and A.J. Greenland. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *J. Exp. Bot.* 58:1231-1244.
- Dunford, R.P., S. Griffiths, V. Christodoulou, and D.A. Laurie. 2005. Characterization of a barley (*Hordeum vulgare* L.) homologue of the *Arabidopsis* flowering time regulator *Gigantea*. *Theor. Appl. Gen.* 110:925-931.
- Ehrenreich, I.M., and M.D. Purugganan. 2006. The molecular genetic basis of plant adaptation. *Am. J. Bot.* 93:953-962.
- Enjalbert, J., I. Goldringer, S. Paillard, and P. Brabant 1999. Molecular markers to study genetic drift and selection in wheat populations. *J. Exp. Bot.* 50:283-290.
- Evanno, G., E. Castella, and J. Goudet. 2006. Evolutionary aspects of population structure for molecular and quantitative traits in the freshwater snail *Radix balthica*. *J. Evol. Biol.* 19(4):1071-1082.

- Flint, J., J. Bond, D.C. Rees, A.J. Boyce, J.M. Roberts-Thomson, L. Excoffier, J.B. Clegg, M.A. Beaumont, R.A. Nichols, and R.M. Harding. 1999. Minisatellite mutational processes reduce  $F_{ST}$  estimates. *Hum. Gen.* 105 :567-576.
- Frankham, R. 1995. Effective population-size adult-population size ratios in wild life: a review. *Genet. Res.* 66:95-107.
- Fu, D.L., P. Szucs, L.L. Yan, M. Helguera, J.S. Skinner, J. von Zitzewitz, P.M. Hayes, and J. Dubcovsky. 2005. Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol. Gen. Gen.* 273:54-65.
- Goldringer, I., J. Enjalbert, A.L. Raquin, and P. Brabant. 2001. Strong selection in wheat populations during ten generations of dynamic management. *Gen. Sel. Evol.* 33:S441-S463.
- Goldringer, I., and T. Bataillon. 2004. On the distribution of temporal variations in allele frequencies, Consequences for the estimation of effective population size and the detection of loci undergoing selection. *Genet.* 168:563-568.
- Goldringer, I., C. Prouin, M. Rousset, N. Galic, and I. Bonnin. 2006. Rapid differentiation of experimental populations of wheat for heading-time in response to local climatic conditions. *Ann. Bot.* 98:805-817.
- Goudet, J. 2002. FSTAT version 2.9.3.2. A program to estimate and test gene diversities and fixation indices. Institut of Ecology, Lausanne, Switzerland. Available at [www.unil.ch/izea/software/fstat.html](http://www.unil.ch/izea/software/fstat.html).
- Goudet, J., and L. Büchi. 2006. The effect of dominance, regular inbreeding and sampling design on  $Q_{ST}$ , an estimator of population differentiation for quantitative traits. *Genet.* 172:1337-1347.
- Goudet, J., and G. Martin. 2007. Under neutrality,  $Q_{ST} \leq F_{ST}$  when there is dominance in an island model. *Genet.* 176:1371-1374.
- Griffiths, S., R.P. Dunford, G. Coupland, and D.A. Laurie. 2003. The evolution of Constans-Like gene families in barley, rice and *Arabidopsis*. *Plant Physiol.* 131:1855-1867.
- Guyomarc'h, H., P. Sourdille, G. Charmet, K.J. Edwards, and M. Bernard. 2002. Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. *Theor. Appl. Gen.* 104:1164-1172.
- Hall, D., V. Luquez, V.M. Garcia, K.R. St Onge, S. Jansson, and P.K. Ingvarsson. 2007. Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes and phenotypic traits. *Evol.* 61(12):2849-2860.
- Kayser, M., S. Brauer, and M. Stoneking. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.* 20:893-900.
- Latta, R.G. 1998. Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptative traits. *Am. Nat.* 151:283-292.
- Le Corre, V., and A. Kremer. 2003. Genetic variability at neutral markers, quantitative trait loci and trait in subdivided population under selection. *Genet.* 164:1205-1219.
- Le Corre, V. 2005. Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol. Ecol.* 14:4181-4192.
- Lee, I., M.J. Aukerman, S.L. Gore, K.N. Lohman, S.D. Michaels, L.M. Weaver, M.C. John, K.A. Feldmann, and R.M. Amasino. 1994. Isolation of *luminidependens*: a gene involved in the control of flowering time in *Arabidopsis*. *Plant Cell* 6:75-83.
- López-Fanjul, C., A. Fernández, and M.A. Toro. 2003. The effect of neutral nonadditive gene action on the quantitative index of population divergence. *Genet.* 164:1627-1633.
- López-Fanjul, C., A. Fernández, and M.A. Toro. 2007. The effect of dominance on the use of the  $Q_{ST} - F_{ST}$  contrast to detect natural selection on quantitative traits. *Genet.* 176:725-727.

- Loskutov, I.G. 2001. Influence of vernalization and photoperiod to the vegetation period of wild species of oats (*Avena spp.*). *Euphyt.* 117:125-131.
- Lynch, M., M. Pfrender, K. Spitze, N. Lehman, J. Hicks, D. Allen, L. Latta, M. Ottene, F. Bogue, and J. Colbourne. 1999. The Quantitative and Molecular Genetic Architecture of a Subdivided Species. *Evol.* 53:100-110.
- Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.* 23:23–35.
- Merilä, J., and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14:892-903.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei M., and F. Tajima. 1981. Genetic drift and estimation of effective population size. *Gen.* 98:625-640.
- O’Hara, R.B., and J. Merilä. 2005. Bias and precision in  $Q_{ST}$  estimates: problems and some solutions. *Genet.* 171:1331-1339.
- Orr, H.A. 2005. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6:119-127.
- Oetjen, K., and H. Reusch. 2007. Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Mol. Ecol.* 24:5156–5157.
- Paillard, S., I. Goldringer, J. Enjalbert, M. Trottet, J. David, C. De Vallaveille-Pope, and P. Brabant. 2000. Evolution of resistance against powdery mildew in winter wheat populations conducted under dynamic management. II- Adult plant resistance. *Theor. Appl. Genet.* 101:457-462.
- Porcher, E., T. Giraud, and C. Lavigne. 2006. Genetic differentiation of neutral markers and quantitative traits in predominantly selfing metapopulations: confronting theory and experiments with *Arabidopsis thaliana*. *Genet. Res.* 87:1-12.
- Putterill, J., R. Laurie, and R. Macknight. 2004. It's time to flower: the genetic control of flowering time. *BioEssays* 2:363-373.
- Raquin, A.L., P. Brabant, B. Rhoné, F. Balfourier, P. Leroy, and I. Goldringer. 2008. Soft selective sweep near a gene that increases plant height in wheat. *Mol. Ecol.* 17:741-756.
- Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Her.* 86:248-249.
- Remington, D.L., and M.D. Purugganan. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. *Int. J. Plant Sc.* 164:S7-S20.
- Rhoné, B., A.L. Raquin, and I. Goldringer. 2007. Strong linkage disequilibrium near the selected Yr17 resistance gene in a wheat experimental population. *Theor. Appl. Gen.* 114:787-802.
- Rhoné, B., C. Remoué, N. Galic, I. Goldringer, and I. Bonnin. 2008. Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. *Mol. Ecol.* 17:930-943.
- Röder, M., V. Korzun, K. Wendehake, J. Plaschke, M.H. Tixier, P. Leroy, and M. Ganal. 1998. A microsatellite map of wheat. *Genet.* 149:2007-2023.
- Rogers, S.M., and L. Bernatchez. 2005. Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Mol. Ecol.* 14:351-361.
- Roux, F., P. Touzet, J. Cuguen, and V. Le Corre. 2006. How to be early flowering: an evolutionary perspective. *Tr. Plant Sc.* 11:375-381.
- Sandring, S., M.A. Riihimäki, O. Savolainen, and J. Ågren. 2007. Selection on flowering time and floral display in an alpine and a lowland population of *Arabidopsis lyrata*. *J. Evol. Biol.* 20:558-567.
- SAS Institute. 2000. *SAS/STAT User Guide, Version 8*. SAS Publishing, Cary, NC.

- Schlötterer, C. 2002. Towards a molecular characterization of adaptation in local populations. *Curr. Op. Gen.Dev.* 12:683-687.
- Sherman, J.D., L. Yan, L. Talbert, and J. Dubcovsky. 2004. A PCR marker for growth habit in common wheat based on allelic variation at the *VRN-A1* gene. *Crop Sc.* 44:1832-1838.
- Siol, M., I. Bonnin, I. Olivieri, J.M. Prospero, and J. Ronfort. 2007. Effective population size associated with self-fertilization: lessons from temporal changes in allele frequencies in the selfing annual *Medicago truncatula*. *J. Evol. Biol.* 20:2349-2360.
- Sourdille, P., S. Singh, T. Cadalen, G.L. Brown-Guedira, G. Gay, L. Qi, B.S. Gill, P. Dufour, A. Murigneux, and M. Bernard. 2004. Microsatellite-based deletion mapping system for the establishment of genetic map – physical map relationships in wheat. *Func. Int. Gen.* 4:12-25.
- Spitze, K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genet.* 135:367-374.
- Staden, R., K.F. Beal, and J.K. Bonfield. 2000. The Staden package. *Met. Mol. Biol.* 132:115-130.
- Stelmakh, A.F. 1987. Growth habit in common wheat (*Triticum aestivum* L. em. Thell.). *Euphyt.* 36:483-488.
- Stinchcombe, J., C. Weinig, M. Ungerer, K.M. Olsen, C. Mays, S.S. Halldorsdottir, M.D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proc. Natl. Acad. Sci.* 101:4712-4717.
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, 100, 158-170.
- Storz, J.F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol. Ecol.* 14:671-688.
- Teshima, K.M., G. Coop, and M. Przeworski. 2006. How reliable are empirical genomic scans for selective sweeps? *Gen. Res.* 16:702-712.
- Vasemägi, A., and C.R. Primmer. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol. Ecol.* 14:3623–3642.
- Vitalis, R., K. Dawson, and P. Boursot. 2001. Interpretation of variation across marker loci as evidence of selection. *Genet.* 158:1811-1823.
- Vitalis, R., K. Dawson, P. Boursot, and K. Belkhir. 2003. DETSEL 1.0: a computer program to detect markers responding to selection. *J. Her.* 94:429-31.
- Waples, R.S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genet.* 121:379-391.
- Weir, B.S., and C.C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. *Evol.* 38:1358-1370.
- Weir, B.S. and Hill. 2002. Estimating *F*-statistics. *Ann. Rev. Genet.* 36:721-50.
- Whitlock, M.C. 1999. Neutral additive genetic variance in a metapopulation. *Genet. Res.* 74:215-221.
- Wilding, C.S., R.K. Butlin, and J. Grahame. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* 14:611-619.
- Worland, A.J., V. Korzun, M.S. Röder, M.W. Ganal, and C.N. Law. 1998. Genetic analysis of the dwarfing gene *Rht8* in wheat. Part II. The distribution and adaptive significance of allelic variants at the *Rht8* locus of wheat as revealed by microsatellite screening. *Theor. Appl. Genet.* 96:1110-1120.

- Yan, L., M. Helguera, K. Kato, S. Fukuyama, J. Sherman, and J. Dubcovsky. 2004. Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theor. Appl. Genet.* 109:1677-1686.
- Yan, L., D. Fu, C. Li, A. Blechl, G. Tranquilli, M. Bonafede, A. Sanchez, M. Valarik, S. Yasuda, and J. Dubcovsky. 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc. Natl. Acad. Sci.* 103:19581-19586.

**Supplementary material 1:** Genome specific primers and PCR conditions for amplifications.

PCR reactions were realised in a final volume of 25 µl containing: 100 µmol/L of each dNTP, 0.5 µmol/L of each primer, 0.5 U of Taq polymerase (Qiagen, Valencia, Calif.), 1X Qiagen buffer, 1X Taq polymerase buffer and 25 ng of genomic DNA (protocol A) or 50 ng of genomic DNA (protocol B).

Cycling consisted of a touch-down profile N°1 as follows: 1 cycle at 94°C for 4 min; 10 cycles at 94°C for 1 min, decreasing annealing temperature by 1°C per cycle starting from 68°C to 59°C for 1 min each temperature, and 72°C for 2.5 min; 21 cycles at 94°C for 1 min, 59°C for 1 min, and 72°C for 2.5 min; a final extension at 72°C for 4 min. For profile N°2, annealing temperature decreased from 66°C to 56°C with 21 cycles using 56°C as annealing temperature.

Gene	Primer name	Primer sequence	Amplification protocol	Touch-down profile	Amplification Size (bases)
LD genome A	LD-A-556U	CGTCAAAGTCCATTTCCGAATC	protocol B	profile N°1	2000
	LD-A-1721L	GACCTGGGCGTGTGTCATTT			
LD genome B	LD-B-L	ACCCAAGCCCATCAACTTCTAT	protocol A	profile N°2	2000
	LD-B-625U	GTGTCAGGTAATATGTATTGCTGC			
LD genome D	LD-D-700U	CACTCATTGCTCCGTATGTA	protocol A	profile N°2	2000
	LD-D-2170L	GTCTGATAATCCAGAAATCCTG			
GI genome A	Gib-A-6U	CCTCACACATGAAAAATGGTTTG	protocol A	profile N°1	1300
	Gif-A-349L	GTGAGGAAGGCGAGGGTCAT			
GI genome B	Gib-B-3U	CAGCCTCACACATGAAAAATG	protocol A	profile N°1	1300
	Gif-B-351L	ATTAGAATGTCACTGAAGCATTGA			
GI genome D	Gib-D-55U	AGAGGCTAGGCGAAATGGTAAC	protocol A	profile N°1	1300
	Gif-D-278L	CTTGCCAGTTCACCGTGTTT			
CO genome A	Co-A-126U	CACGGCGGAGATGTCGTAC	protocol A	profile N°1	300
	Co-A-390L	TGGAAATAATAAGATGAGAGACAAGTAG			
CO genome B	Co-B-141U	GTATGATGCACAAAATTCTAACCA	protocol A	profile N°2	300
	Co-B-377L	TGAGACAAGTACGATAAGATGGG			
CO genome D	Co-D-121U	GGCTTCACGGCGGAGATGC	protocol A	profile N°1	300
	Co-D-541L	CCCATGAACTGCGGCGACAT			

## Supplementary material 2 : Protocol for competitive allele-specific PCR.

Initial genome-specific PCR amplifications of the genes were carried out as described before. PCR products were then diluted 1/50 and 2  $\mu$ L were used for subsequent SNP genotyping by competitive allele-specific PCR. Reactions were carried out in 25  $\mu$ L total volume, with 3 mM MgCl<sub>2</sub>, 0.6 mM of each dNTP, 1U Taq polymerase (Qiagen), Taq polymerase buffer (Qiagen), and the three following primers : (1) 0.5  $\mu$ M of a (short) forward primer designed from the sequence immediately upstream to the SNP, with the base at its 3' extremity corresponding to the SNP position, (2) 0.25  $\mu$ M of a (long) similar forward primer, but with the base at the 3' extremity corresponding to the alternate allele of the SNP and with a 5' tail added, so that PCR products originating from this long forward primer are 19 bp longer than those originating from the short forward primer, and (3) 0.5 $\mu$ M of a reverse primer designed to get an amplified product with either the short or the long forward primer. The competition between both forward primers results in a much higher priming efficiency for the primer with a correct match at the 3' base, so that depending on the allele at the SNP position, either the short or the long forward primer is preferentially involved in the product. Thermal cycling conditions were : 94° 5 min, 35 cycles: ( 94° 30 sec, 65° or 60° 30 sec, 72° 30 sec), final extension 72° 5 min. The products were electrophoresed on 4% MetaPhor<sup>TM</sup> (TEBU, Le Perray, France) agarose gel with 1mg/mL ethidium bromide during 4 hours at 120V, so that the 19 bp size differences between the two SNP alleles were easy to read.

	Long forward primers (5' to 3')	Short forward primers (5' to 3')	Reverse primers (5' to 3')
LD	CACGACGTTGTAACGACTTGTCATCTTGTGTACGAATT,	TTGTCATCTTGTGTACGAATTC	CTGGAAGAACACAAATAATGACG
GI	CACGACGTTGTAACGACCAGAAGAGGCTAGGCGAAAT	CAGAAGAGGCTAGGCGAAAC	AGCCTCGATAACCCCAT
CO	CACGACGTTGTAACGACCAAGCACTCTTGGTCGATCTT	CAAGCACTCTTGGTCGATCTC	AATTCTATGGACTTGATCTCTAT

Genotyping results on these three genes were completed and checked by using a variant of competitive allele-specific PCR : here, a 5' tail containing a restriction site was added to the short forward primer, so that both forward primers had the same size. Competition bias due to primer size differences was then reduced. Long forward primers and reverse primers were the same as above. Competitive allele-specific PCRs were then carried out as described above, except for the reaction volume which was 22  $\mu$ l. These products were then entirely digested with 3 units of restriction enzyme in a total volume of 27  $\mu$ l at 37°C overnight, then electrophoresed as described above. For *LD* and *GI*, the restriction enzyme used was EcoRV and the 5' tail for short forward primers was (5' to 3') CACGACGTTGTAAGATATC. For *CO*, the restriction enzyme used was EcoRI and the 5' tail for short forward primers was (5' to 3') CACGACGTTGTAAGAATTC.

## V-4 Conclusions

Le suivi de l'évolution de la précocité en relation avec certains gènes impliqués montre que l'évolution du caractère au cours des générations dans les différentes populations s'accompagne de changements dans la composition allélique de ces gènes qui peuvent être importants en particulier pour les gènes majeurs tels que *VRN-1*. Cependant pour d'autres gènes à effets plus faibles, la réponse à la sélection est caractérisée par la sélection de combinaisons d'allèles à différents locus induisant la mise en place de déséquilibres de liaison entre gènes, et de faibles changements de fréquences alléliques aux locus pris individuellement.

Le cas du gène *VRN-1* est intéressant car bien qu'il explique plus de 80% de la variation du caractère de précocité en conditions non vernalisantes, son effet sur la variation de la précocité au champ est bien plus modéré rappelant que ce caractère est complexe et résulte de l'interaction d'un grand nombre de gènes dont l'expression est induite par des signaux environnementaux et intrinsèques au cours du développement.

Une fois encore, la population de Vervins se démarque des autres populations avec des changements de fréquences alléliques importants dès les premières générations liés vraisemblablement au goulot d'étranglement subi à l'installation de la population et mis en évidence par l'étude des marqueurs microsatellites. Ces variations très importantes profitent également à des polymorphismes initialement non trouvés dans la population parentale (polymorphisme aux locus *VRN-1Apr* et *FT-Aint1*) qui apparaissent par mutation ou par migration et qui augmentent fortement en fréquence dans cette population. Les déséquilibres de liaison très forts trouvés dans cette population en génération 12 entre zones polymorphes des gènes sont à mettre en relation avec les forts déséquilibres de liaison déjà observés entre les locus microsatellites. Cette situation est d'ailleurs très problématique pour détecter l'implication de certaines zones polymorphes dans la variation du caractère et leur potentielle sélection. Si les populations expérimentales de blé offrent un support exceptionnel pour étudier l'impact de la sélection d'un caractère adaptatif sur l'évolution des gènes impliqués, il est clair que l'association des polymorphismes avec le caractère doit également être étudiée sur un matériel pour lequel on peut s'affranchir de la structuration génétique intra population tel que des populations recombinantes allogames ou des collections à base génétique large utilisées en génétique d'association.

## *Chapitre 6*

### *CONCLUSIONS ET PERSPECTIVES*

## V-4 Conclusions

Le suivi de l'évolution de la précocité en relation avec certains gènes impliqués montre que l'évolution du caractère au cours des générations dans les différentes populations s'accompagne de changements dans la composition allélique de ces gènes qui peuvent être importants en particulier pour les gènes majeurs tels que *VRN-1*. Cependant pour d'autres gènes à effets plus faibles, la réponse à la sélection est caractérisée par la sélection de combinaisons d'allèles à différents locus induisant la mise en place de déséquilibres de liaison entre gènes, et de faibles changements de fréquences alléliques aux locus pris individuellement.

Le cas du gène *VRN-1* est intéressant car bien qu'il explique plus de 80% de la variation du caractère de précocité en conditions non vernalisantes, son effet sur la variation de la précocité au champ est bien plus modéré rappelant que ce caractère est complexe et résulte de l'interaction d'un grand nombre de gènes dont l'expression est induite par des signaux environnementaux et intrinsèques au cours du développement.

Une fois encore, la population de Vervins se démarque des autres populations avec des changements de fréquences alléliques importants dès les premières générations liés vraisemblablement au goulot d'étranglement subi à l'installation de la population et mis en évidence par l'étude des marqueurs microsatellites. Ces variations très importantes profitent également à des polymorphismes initialement non trouvés dans la population parentale (polymorphisme aux locus *VRN-1Apr* et *FT-Aint1*) qui apparaissent par mutation ou par migration et qui augmentent fortement en fréquence dans cette population. Les déséquilibres de liaison très forts trouvés dans cette population en génération 12 entre zones polymorphes des gènes sont à mettre en relation avec les forts déséquilibres de liaison déjà observés entre les locus microsatellites. Cette situation est d'ailleurs très problématique pour détecter l'implication de certaines zones polymorphes dans la variation du caractère et leur potentielle sélection. Si les populations expérimentales de blé offrent un support exceptionnel pour étudier l'impact de la sélection d'un caractère adaptatif sur l'évolution des gènes impliqués, il est clair que l'association des polymorphismes avec le caractère doit également être étudiée sur un matériel pour lequel on peut s'affranchir de la structuration génétique intra population tel que des populations recombinantes allogames ou des collections à base génétique large utilisées en génétique d'association.

## **Chapitre VI : Conclusions et perspectives**

### **VI-1 Etude de l'évolution conjointe des marqueurs microsatellites, de la précocité de floraison et des gènes candidats : synthèse des principaux résultats**

Le suivi de l'évolution de marqueurs microsatellites, de gènes et de caractères liés à la valeur sélective dans 3 populations de gestion dynamique évoluant dans des environnements contrastés a montré que ces populations évoluent au cours des générations et se différencient de façon significative pour ces 3 niveaux de diversité. La précocité de floraison présente une différenciation bien plus élevée que les marqueurs microsatellites dont les variations inter populations fournissent une image du niveau de différenciation moyen de l'ensemble du génome. Ceci traduit le fait que la précocité de floraison est sélectionnée. Cependant cette sélection ne se fait pas de façon directe et semble être le fait de la sélection agissant sur des caractères corrélés eux-mêmes sélectionnés. La sélection pour la précocité peut être mise en relation avec les caractéristiques climatiques des sites de culture, les populations du Nord fleurissant plus tardivement que les populations du Sud. Différentes stratégies de sensibilité aux facteurs environnementaux se sont mise en place dans les populations permettant un ajustement de la floraison en fonction des conditions environnementales de chaque site. Ainsi la population du Moulon développe plutôt une stratégie de sensibilité aux températures froides permettant à la plante d'initier la floraison une fois la saison hivernale passée et alors que la longueur du jour augmente au printemps. Au contraire, dans la population de Toulouse, les génotypes les plus précoces s'affranchissent de la sensibilité à tous les signaux environnementaux.

La différenciation importante de la précocité, en relation avec l'adaptation locale des populations, s'accompagne d'une différenciation importante au gène *VRN-1*, gène majeur de la voie de la vernalisation, et montré comme clairement associé à la précocité. En revanche d'autres gènes associés à la précocité de floraison évoluent de façon moins importante de sorte que la sélection à ces locus ne peut être invoquée.

### Cas particulier de la population de Vervins

L'étude conjointe des différents niveaux de diversité a montré que la population de Vervins a subi une évolution particulière comparée à celles des autres populations. Dès la première génération étudiée, le très faible effectif efficace laisse penser que la population a subi un fort goulot d'étranglement. La reconstruction d'haplotypes multilocus vient renforcer cette hypothèse puisque un tiers des génotypes présentent la même combinaison d'allèles à tous les locus et suggère donc qu'une lignée majoritaire domine la population en génération 2. Toutefois, la richesse allélique n'en est que peu affectée, tant au niveau des marqueurs microsatellites qu'au niveau des gènes eux-mêmes et les caractères ne montrent pas de réduction drastique de leur niveau de variabilité. Au cours des générations suivantes, les fréquences alléliques des gènes et des marqueurs microsatellites évoluent de façon importante du fait de la dérive et de la sélection. Ceci s'accompagne de déséquilibres de liaison importants qui se renforcent au cours des générations. Dans ce contexte de forte dérive, une baisse de l'efficacité de la sélection à certains locus serait à craindre. Cependant, il semble que ce ne soit pas le cas puisque cette population évolue vers une plus grande tardivité de façon continue au cours des générations en réponse aux conditions climatiques locales. Elle ne semble pas non plus souffrir d'une plus faible productivité que les autres populations. Au contraire, c'est cette population qui présente les plus fortes valeurs de poids de 1000 grains en génération 12 quelles que soient les conditions de vernalisation (4S, 8S et champs), ainsi qu'une plus forte valeur moyenne de la hauteur. Cette évolution particulière à Vervins entraîne même une différenciation génétique des populations pour la hauteur en génération 12 jamais observée au préalable dans les études des populations de gestion dynamique. Enfin des allèles initialement non présents dans les lignées parentales et arrivant par mutation et par migration sont sélectionnés et observés en forte fréquence en génération 12 alors que la force de la dérive devrait plutôt tendre à les éliminer.

## **VI-2 Apports et limites de l'étude dans la compréhension des bases génétiques de l'adaptation**

### **VI-2.1 Les apports de l'étude**

Bien que cette étude des bases génétiques de l'adaptation de la précocité dans les populations expérimentales de blé ne soit qu'un exemple et qu'elle se base sur un nombre limité de gènes, elle apporte avancées intéressantes dans la compréhension des réponses génétiques mises en jeu lors de l'adaptation et dont les principales interrogations sont présentées en introduction :

#### **VI-2.1.1 Effets des gènes sélectionnés**

La sélection de gènes majeurs, tels que le gène *VRN-1*, permet une évolution rapide des populations dès les premières générations. Ce résultat illustre le fait que la sélection peut jouer sur la variation de gènes à effets majeurs sur les caractères contrairement à ce qui fut longtemps admis (modèle infinitésimal de Fisher) et va dans le sens d'autres observations expérimentales résumées dans Orr (2005). Cependant la précocité continue d'évoluer de façon importante alors que les polymorphismes identifiés au gène *VRN-1* n'évoluent plus de façon significative entre la génération 7 et la génération 12. Cette observation, combinée au fait que d'autres gènes à effets plus faibles continuent d'évoluer, suggère qu'à l'approche de l'optimum, la sélection opère sur des gènes à effets faibles permettant un ajustement plus fin du phénotype.

#### **VI-2.1.2 Standing genetic variation vs. mutation**

L'utilisation de croisements multi-parentaux pour constituer les populations de gestion dynamique a permis l'action rapide de la sélection grâce à la disponibilité d'une certaine diversité initiale. Bien que nous n'ayons pas pu clairement trancher entre mutation et migration, la nouvelle diversité apparue aux cours des générations semble jouer un rôle important dans la réponse à la sélection dans la population de Vervins qui a subi un goulot

d'étranglement dans les premières générations. Ces résultats suggèrent que l'adaptation locale des populations est principalement le fait de la sélection agissant sur la diversité disponible. Mais, lorsque la diversité est réduite, l'apparition d'une nouvelle diversité joue alors un rôle important pouvant conduire à des évolutions très importantes des fréquences d'allèles initialement présents en très faibles proportions et qui n'ont pas été perdus par dérive (par exemple pour le polymorphisme de l'intron 1 du gène *FT* non polymorphe dans les lignées parentales, cf. Chapitre V). Pour approfondir cette question il pourrait être intéressant d'étudier l'évolution d'un plus grand nombre de gènes de floraison initialement non polymorphes dans la population parentale et qui ne sont pas génotypés d'habitude.

### **VI-2.1.3 Pléiotropie**

Les variations de gènes (*Ppd* et *FT*), pourtant polymorphes parmi les lignées parentales et dans la population initiale et montrés comme impliqués dans les variations de la précocité de floraison dans la littérature (Beales *et al.* 2007 ; Bonnin *et al.* 2008), ne montrent pas de trace de sélection entre les populations. Bien que la faible puissance des méthodes de détection de la sélection utilisées puisse être invoquée, cette absence de signal pourrait aussi être due à des pressions de sélection contradictoires s'exerçant sur ces gènes du fait de leur implication dans l'architecture d'autres caractères. Pour le gène *Ppd*, cette hypothèse est à considérer sérieusement du fait de son association possible avec la variation de la hauteur (Börner *et al.* 1993) qui est également sélectionnée de façon unidirectionnelle dans les populations.

### **VI-2.1.4 Rôle de l'épistasie**

Les études d'association entre les gènes et le caractère montrent clairement que les interactions entre gènes jouent un rôle important dans l'architecture du caractère de floraison chez le blé. Ceci n'est pas vraiment surprenant si l'on considère l'architecture de la précocité de floraison par exemple chez *A. Thaliana* où un grand nombre de gènes interagissent en cascades dans différentes voies de régulation alors que d'autres intègrent l'information des ces différentes voies pour initier la floraison. Dans un tel contexte, il semble assez intuitif que la mutation d'un gène en amont de la cascade de régulation ait une incidence sur l'expression des gènes situés plus en aval. Du fait de la nature polyploïde du génome du blé, les interactions observées dans notre étude sont de 2 ordres : entre copies d'un même gène ou

entre gènes eux-mêmes. Cette situation complique l'analyse de l'effet des interactions sur la variation du caractère car certaines combinaisons d'allèles multi-locus sont peu ou pas représentées, réduisant alors la puissance statistique de l'analyse de variance. De plus la fonction des gènes identifiés par l'homologie de leur séquence avec celles de gènes d'*A. Thaliana* n'est pas connue chez le blé et la compréhension des mécanismes impliqués dans l'interaction entre gènes nécessiterait des études précises de leur positionnement dans la chaîne de régulation. En revanche la relation d'épistasie trouvée entre les copies du gène *VRN-1* est bien documentée et joue un rôle important dans la détermination du caractère H/P puisque la présence d'un allèle de type « P » à l'une des 3 copies du gène est suffisante pour que le phénotype soit de type Printemps. Du fait de cette interaction, la sélection pour le type Hiver nécessite la sélection combinée d'allèles associés au type H à chaque locus créant ainsi des déséquilibres de liaison entre les différentes copies. Ainsi cette étude tend à montrer que l'interaction entre locus favorise l'action de la sélection sur des combinaisons d'allèles multi-locus plutôt que sur des locus individuels.

### **VI-2.1.5 Type de mutations**

Sur les 13 locus polymorphes suivis dans les populations, seuls 4 impliquent des polymorphismes de séquence codante (*VRN-1A<sub>ex7</sub>*, *Ppd-1D<sub>ex8</sub>*, *FT-D<sub>ex3</sub>* et *GI-B<sub>ex5</sub>*). Parmi ces polymorphismes, seul *VRN-1A<sub>ex7</sub>* évolue de façon significative dans les populations mais cette évolution semble due à la sélection de la région promotrice située 12 kb en amont. Pour les autres locus situés soit dans un intron, soit dans un promoteur (*VRN-1A*) pour les gènes publiés (*FT*, *VRN-1* et *Ppd-1*) ce sont principalement les régions non codantes qui contiennent des polymorphismes causaux, même si aucune association avec la précocité n'est détectée pour *Ppd-1* dans les populations (association montrée par ailleurs dans une collection de blé et par étude d'expression, Beales et al. 2007). Dans le cas de *VRN-1*, l'observation de plusieurs allèles dominants altérant la région promotrice de *VRN-1A* ou les introns de *VRN-1B* et *VRN-1D* et conduisant au même phénotype laisse supposer que les mutations dans des régions régulatrices permettent d'obtenir facilement des phénotypes Printemps. Cette observation vient appuyer l'avantage des mutations dans des régions régulatrices sur les mutations des zones codantes lors de la sélection.

## VI-2.2 Les limites

Du fait du nombre limité de gènes analysés, nous ne pouvons apporter de contribution à la question du nombre de gènes participant à l'adaptation pour la précocité dans nos populations. En effet, du fait de la nature polyploïde et malgré l'intérêt agronomique évident du blé tendre, très peu de ressources génomiques sont disponibles à l'heure actuelle, en dehors des polymorphismes nucléotidiques des 2 gènes majeurs (*VRN* et *Ppd*) de la précocité de floraison, identifiés bien avant l'utilisation à grande échelle des marqueurs moléculaires.

C'est pourquoi une démarche d'identification de gènes candidats par recherche des séquences orthologues de gènes du riz ou d'*Arabidopsis thaliana* a été initiée en amont de cette thèse. Cette approche s'est avérée fructueuse pour 2 gènes sur 4 (*LD* et *FT*) pour lesquels l'association avec le caractère a été montrée dans une collection à base génétique large ainsi que dans la présente étude. Pour le fragment de gène (*GI*) qui ne présente pas d'association avec le caractère de précocité dans notre matériel et en génétique d'association sur une collection mondiale de blé tendre, il est possible que cette région orthologue de gènes d'*A. Thaliana* soit un fragment de gènes non fonctionnels. En effet la duplication de gènes est un phénomène souvent observé au sein du génome des organismes végétaux. Cette tendance pourrait être particulièrement importante chez le blé du fait de la nature hexaploïde du génome qui multiplie par 3 le nombre de copies de gène. La théorie prédit que, du fait de la redondance, des mutations délétères peuvent s'accumuler au sein de copie de gène conduisant à terme à la perte de fonction de certaines copies qui évoluent alors de façon neutre (Moore et Purugganan 2005). Ainsi, la recherche de gènes candidats par homologie de séquences chez une espèce polyploïde pourrait conduire à l'identification de régions de gènes non fonctionnelles alors que la mise en œuvre de ce type d'approche est délicate et coûteuse.

Pour répondre à la question du nombre de gènes impliqués dans la réponse à la sélection du caractère, une étude de type « genome scan » pourrait être envisagée avec un grand nombre de marqueurs répartis sur l'ensemble du génome ou des marqueurs positionnés dans les intervalles de confiance des QTL de floraison pour identifier les régions cibles de la sélection (données rendues disponibles par une étude de méta-analyse de cartographie de QTL de gènes de floraison chez le blé parue en 2007, Hannocq et al. 2007). Cependant, notre étude utilisant 3 approches différentes pour détecter les polymorphismes de gènes candidats sous sélection divergente met en évidence que ces méthodes ne permettent pas de détecter tous les gènes associés à la variation du caractère (*FT* et *LD*). Par ailleurs, l'étude du déséquilibre de liaison à partir des locus microsatellites montre que le déséquilibre de liaison s'étend sur une longue

distance le long du chromosome 3B (30 cM) et que les DL entre locus situés sur des chromosomes différents peuvent également être importants. Cette structuration observée à l'intérieur des populations peut être extrêmement handicapante pour détecter les régions du génome sous sélection en utilisant des marqueurs anonymes. Cela ne va d'ailleurs pas sans poser de problèmes dans l'approche que nous avons utilisée puisque les forts déséquilibres de liaison conduisent à une évolution significative du gène *CO* en relation avec la sélection du gène *VRN-1* alors que ces 2 gènes d'un même chromosome sont localisés sur des brins différents. La détection de cette forte structuration intra population montre les limites des études d'associations effectuées dans nos analyses qui, bien que nécessaires, peuvent conduire à des conclusions erronées si une validation de l'implication des gènes identifiés n'est pas effectuée par ailleurs. La mise en œuvre de la génétique d'association, en tenant compte de la structuration d'une collection mondiale de blé, a permis de valider au laboratoire l'implication de certains polymorphismes dans la variation de la précocité de floraison. J'ai par ailleurs initié une étude d'expression des gènes pour tenter de mettre en évidence des patrons d'expression différents entre génotypes (4 lignées parentales des populations de GD de type « Printemps » et de types « Hiver ») présentant des allèles différents aux locus étudiés (gènes *LD* et *FT* en particulier) dans des conditions de vernalisation complète et en l'absence de vernalisation.

### **VI-2.3 Les nouvelles interrogations**

Bien que l'approche employée dans cette étude comporte certaines limitations, les résultats obtenus ouvrent la voie à de nouvelles explorations du mécanisme de l'adaptation :

#### **VI-2.3.1 Allogamie**

L'étude des marqueurs microsatellites a mis en évidence que les populations de GD utilisées dans ce travail sont fortement autogames. L'autogamie a une forte incidence sur l'évolution des populations. En premier lieu elle tend à réduire de façon importante l'effectif efficace des populations, de plus elle conduit à l'établissement de forts déséquilibres de liaison à l'échelle du génome qui peuvent être accrus par les effets combinés de la dérive et de la sélection. L'étude de l'évolution des gènes dans les populations de GD confirme les attendus théoriques selon lesquels, dans le cas autogame, la sélection entraîne une réponse rapide en favorisant les

individus qui cumulent le plus d'allèles avantageux. Cette évolution rapide tend également à créer des déséquilibres de liaison entre gènes et peut conduire à terme à la fixation d'allèles faiblement délétères rendant la sélection moins efficace (« background selection », Charlesworth 2003). Il serait donc intéressant de pouvoir comparer la réponse à la sélection d'un caractère mise en œuvre dans le cas autogame à ce qui se passe chez des espèces allogames. Lors de la mise en place des populations de gestion dynamique, 3 pools ont été distribués sur l'ensemble du réseau dont un pour lequel une des lignées de départ présente un gène récessif de stérilité mâle cytoplasmique (population PS). Dans les populations issues du pool PS ce gène de stérilité mâle a été maintenu artificiellement en semant d'une année sur l'autre uniquement les descendances des plantes mâles stériles, nécessairement issues d'évènements d'allofécondations. Ces populations « allogamisées » artificiellement ont été relativement peu étudiées jusqu'à présent. David (1999) a montré que contrairement aux populations autogames (pool PA et PB) qui se différencient selon un gradient Nord/Sud de précocité, celles-ci ont toutes évolué, au bout de 8 générations, vers une plus grande précocité, et ce quel que soit leur site de culture. Bien que la diversité initiale de ces populations PS (populations issues du croisement de 67 lignées parentales) soit bien supérieure à la diversité initiale des populations issues du pool PA étudiées dans le cadre de ma thèse (16 lignées parentales) et limite les comparaisons, une analyse plus fine de la réponse mise en œuvre pour l'adaptation des populations allogames serait très instructive.

### **VI-2.3.2 Polyploïdie**

Cette étude soulève la question de l'intérêt de la polyploïdie dans la réponse à l'adaptation. Les cas de polyploïdie sont très fréquents dans la nature et certains auteurs se sont intéressés aux conséquences évolutives de la polyploïdie (cf. revues de Comai 2005 et Otto 2007 sur le sujet). Dans le cas du gène *VRN-1*, il semble que la présence de 3 copies fonctionnelles du même gène offre une grande possibilité de réponse à la sélection face à des changements d'environnements conduisant à la sélection de combinaisons alléliques différentes dans les différents milieux. Ceci suggère que les polyploïdes pourraient répondre plus rapidement que les diploïdes aux pressions évolutives dans le cas où la mutation est limitée (Strasburg et Gross 2008).

### **VI-3 Implications pour la conservation dynamique des ressources génétiques du blé**

L'étude conjointe de l'évolution des marqueurs microsatellites, des gènes et des caractères montre qu'une grande diversité est conservée à l'échelle des 3 populations considérées et donc à l'échelle du réseau de populations dans son ensemble. Ceci est possible grâce à la dérive et à la sélection qui conduisent à la différenciation des populations à toutes les échelles de diversité considérées. Mais cette observation doit aussi être comparée à la grande diversité observée à l'intérieur de chaque population et en particulier dans les populations de Toulouse et de Vervins après 12 générations d'évolution. Ce résultat plus surprenant s'explique par le fait que, dans un même environnement, plusieurs combinaisons alléliques différentes peuvent conduire à un même phénotype en particulier dans le contexte polyploïde comme souligné précédemment. De plus cette diversité peut également s'expliquer par les événements de mutations ou de migrations qui créent une nouvelle diversité sur laquelle la sélection peut jouer. Toutefois, après 12 générations d'évolution, il est également possible que les populations n'aient pas encore atteint leur optimum phénotypique. En particulier l'évolution importante observée à tous les niveaux pour la population de Vervins et l'apparition de lignées au cours des générations sont des signaux inquiétants quant à l'évolution future de cette population, si la sélection d'une lignée particulière venait à menacer la diversité jusque là préservée. C'est pourquoi une nouvelle étude a été initiée dans le cadre de la thèse d'Yves Rousselle (2006-2009) sur la quantité et la qualité des migrations à apporter entre les différentes populations pour fournir un nouveau potentiel adaptatif aux populations sans compromettre leur adaptation locale.

## *BIBLIOGRAPHIE*

## Bibliographie

- Alonso-Blanco C, Mendez-Vigo B, Koornneef M (2005) From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *International Journal of Developmental Biology*, 49, 717-732.
- Ardlie KG, Kruglyak L, Seielstad L (2002) Patterns of linkage disequilibrium in the human genome. *Nature Review Genetics*, 3, 299-309.
- Barrett RDH Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*. 23, 38-44.
- Beales J, Turner A, Griffiths S, Snape JW, Laurie DA (2007) A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 115, 1432-2242.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13, 969-980.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France), <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>.
- Black CW, Krafur ES (1985) A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theoretical Applied Genetics*, 70:491-496.
- Bonin A, Bellemain E, Bronkein Eidesen P, Pompanon F, Brchmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, 13, 3261-3273.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology Evolution*, 23, 773-783.
- Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology* 16, 3737-3758.
- Bonnin I, Rousset M, Madur D, Sourdille P, Dupuits C, Brunel D, Goldringer I (2008) *FT* genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. *Theoretical and Applied Genetics*, 116, 383-394
- Börner A, Worland AJ, Plashke J, Shuman E, Law CN (1993) Pleiotropic effects of genes for reduced height (*Rht*) and day-length insensitivity (*Ppd*) on yield and its components for wheat growth in Middle Europe. *Plant Breeding*, 11, 204-216.

- Boudry P, McCombie H, van Dijk H (2002). Vernalization requirement of wild beet *Beta vulgaris* ssp. *maritima*: among population variation and its adaptive significance. *Journal of Ecology*, 90, 693-703.
- Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques*, 24-28.
- Chao S, Zhang W, Dubcovsky J, Sorrells M (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes. *Crop Science*, 47, 1018-1030.
- Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, 70, 155-174.
- Charlesworth D (2003) Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358, 1051-1070.
- Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *Journal of Experimental Botany*, 58, 1231-1244.
- Comai L (2005) The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, 6, 836-846.
- David J (1992) Approche méthodologique d'une gestion dynamique des ressources génétiques chez le Blé tendre (*Triticum aestivum* L.). Thèse, Institut National Agronomique Paris Grignon.
- David JL, Savy Y, Trottet M, Pichon M (1992) Méthode de gestion dynamique de la variabilité génétique. Exemple d'un réseau expérimental de populations composites de blé tendre. In : Complexes d'espèces, Flux de gènes et ressources génétiques (Ed. BRG) 337-350, Lavoisier, Cachan, France.
- David JL, Savy Y, Brabant P (1993) Outcrossing and selfing evolution in populations under directional selection. *Heredity*, 71, 642-651.
- David P, Pujol B, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, 16, 2474-2487.
- Demotes-Mainard S, Doussinault G, Meynard JM (1995) Effects of low radiation and low temperature at meiosis on pollen viability and grain set in wheat. *Agronomie*, 15, 357-365.
- Demuth JP, Wade MJ (2006) Experimental methods for measuring gene interactions. *Annual Review of Ecology, Evolution, and Systematics*, 37, 289-316.

- Ehrenreich IM, Purugganan MD (2006) The molecular genetic basis of plant adaptation. *American Journal of Botany*, 97, 953-962.
- Endler JA (1986) *Natural selection in the wild*. Princeton University Press, Princeton.
- Enjalbert J, Goldringer I, David J, Brabant P (1998) The relevance of outcrossing for the dynamic management of genetic resources in predominantly selfing *Triticum aestivum* L. (bread wheat). *Génétique, Sélection, Evolution*, 30: S197-S211.
- Enjalbert J, Goldringer I, Paillard S, Brabant P (1999) Molecular markers to study genetic drift and selection in wheat populations. *Journal of experimental botany*, 50, 283-290.
- Enjalbert J (1999) Diversité et régime de reproduction de populations de blé tendre (*Triticum aestivum* L.) menées en gestion dynamique. Thèse, Institut National Agronomique Paris Grignon.
- Erickson DL, Fenster CB, Stenoien HK, Price D (2004) Quantitative trait analyses and the study of evolutionary process. *Molecular Ecology*, 13, 2505-2522.
- Excoffier, L. G. Laval, and S. Schneider (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
- Fay JC, Wittkopp PJ (2008) Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*, 100, 191-199.
- Fisher RA (1930) *The genetical theory of natural selection*. Oxford University Press, Oxford.
- Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews Genetics*, 2, 437-445.
- Fox CW, Roff DA, Fairbairn DJ (2001) *Evolutionary ecology: concepts and case studies*. Oxford University Press, Oxford.
- Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, 66, 95-107.
- Franks SJ, Sim S, Weis AE (2007) Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceeding of the National Academy of Sciences*, 104, 1278-1282.
- Fu DL, Szucs P, Yan LL et al. (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Molecular Genetics and Genomics*, 273, 54-65.
- Gardner KM, Latta RG (2007) Shared quantitative trait loci underlying the genetic correlation between continuous traits. *Molecular Ecology*, 16, 4195-4209.
- Garnier-Gere P, Dillmann C (1992) A computer program for testing pairwise linkage disequilibrium in subdivided populations. *Journal of Heredity*, 83, 239.

- Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceeding of the Royal Society B*, 273, 3011-3019.
- Goldringer I, Enjalbert J, Raquin AL, Brabant P (2001) Strong selection in wheat populations during ten generations of dynamic management, *Génétique, Sélection, Evolution*, 33, S441-S463.
- Goldringer I, Bataillon T (2004) On the distribution of temporal variations in allele frequencies, Consequences for the estimation of effective population size and the detection of loci undergoing selection. *Genetics*, 168, 563-568.
- Goldringer I, Prouin C, Rousset M, Galic N, Bonnin I (2006) Rapid differentiation of Experimental populations of wheat for heading time in response to local climatic conditions. *Annals of Botany*, 98, 805-817.
- Goudet J (2002) FSTAT version 2.9.3.2. A program to estimate and test gene diversities and fixation indices. Institut of Ecology, Lausanne, Switzerland, Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995).
- Goudet J, Büchi L (2006) The effect of dominance, regular inbreeding and sampling design on  $Q_{st}$ , an estimator of population differentiation for quantitative traits. *Genetics*, 172, 1337-1347.
- Goudet J, Martin G (2007) Under neutrality,  $Q_{st} \leq F_{st}$  when there is dominance in an island model. *Genetics*, 176, 1371-1374.
- Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The Evolution of CONSTANS-Like Gene Families in Barley, Rice, and Arabidopsis. *Plant Physiology*, 131, 1855–1867.
- Guyomarc'h H, Sourdille P, Charret G, Edwards KJ, Bernard M (2002) Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. *Theoretical Applied Genetics*, 104, 1164-1172.
- Hanocq E, Laperche A, Jaminon O, Lainé AL, Legouis J (2007) Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. *Theoretical and Applied Genetics*, 114, 569-584.
- Hansen TF (2006) The evolution of genetic architecture. *Annual Review of Ecology, Evolution, and Systematics*, 37, 123-57.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S, David J (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution*, 24, 1506-1517.
- Hedrick PW (2005) *Genetics of populations*. 3ème édition. Jones and Batlett Publishers, Sudbury.
- Hendry AP, Kinnison MT (2001) An introduction to microevolution: rate, pattern, process. *Genetica*, 112-113, 1-8.

- Henry JP, Pontis C, David JL, Gouon PH (1991). An experiment on dynamic conservation of genetic resources with metapopulations, In : Seitz A. and Loeschcke V. (eds), Species Conservation : a Population Biological Approach, Basel, 185-198.
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, 61, 995-1016.
- Hoffman JJ, Amos W (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14, 599–612
- Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics*, 130, 195-204.
- Kauer M, Zangerl B, Dieringer D, Schlötterer C (2002) Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. *Genetics*, 160, 247-256.
- Kingsolver JG, Hoekstra HE, Berrigan D, Vignieri SN, Hoang A, Gibert P, Beerli P (2001) The strength of phenotypic selection in natural populations. *The American Naturalist*, 157, 245-261.
- Kingsolver JG, Pfennig DW (2007) Patterns and Power of Phenotypic Selection in Nature, *BioScience*, 57, 561-572.
- Kingsolver JG, Massie KR, Shlichta JG, Smith MH, Ragland GJ, Gomulkiewicz R (2007) Relating environmental variation to selection on reaction norms: an experimental test. *The American Naturalist*, 169, 163-174.
- Kuchel H, Hollamby G, Landridge P, Williams K, Jefferies SP (2006) Identification of genetic loci associated with ear-emergence in bread wheat. *Theoretical and Applied Genetics* 113:1103-1112
- Kraakman A, Niks R, Van den Berg P, Stam P, Van Eeuwijk P (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, 168, 435-446.
- Lande R (1979) Quantitative genetics analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, 33, 402-416.
- Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, 37, 1210-1226.
- Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *The American Naturalist*, 151, 283-292.
- Le Boulc'h V., David J., Brabant P., de Vallavieille-Pope C., 1994. Dynamic conservation of variability : responses of wheat populations in different selective forces including powdery mildew, *Genetics Selection Evolution*, 26, S221-S240.

- Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in subdivided population under selection. *Genetics*, 164, 1205-1219.
- Le Corre V (2005) Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Molecular Ecology*, 14, 4181–4192
- Leinonen T, O’Hara RB, Cano JM, Merilä J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, 21, 1-17.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74, 175-195.
- López-Fanjul C, Fernández A, Toro MA (2003) The effect of neutral nonadditive gene action on the quantitative index of population divergence. *Genetics*, 164, 1627-1633.
- López-Fanjul C, Fernández A, Toro MA (2007) The effect of dominance on the use of the  $Q_{st}$ - $F_{st}$  contrast to detect natural selection on quantitative traits. *Genetics*, 176, 725-727.
- Loskutov IG (2001) Influence of vernalization and photoperiod to the vegetation period of wild species of oats (*Avena* spp.). *Euphytica*, 117, 125–131.
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, Massachusetts.
- Mackay TFC (2001) The genetic architecture of quantitative traits. *Annual review of genetics*, 35, 303-339.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetical Research*, 23, 23–35
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution*, 17, 285-291.
- Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, 14, 892-903.
- Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits ? *Nature Reviews Genetics*, 8, 845-856.
- Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology*, 8, 122-128
- Mouradov A, Cremer F, Coupland G (2002) Control of flowering time, Interacting pathways as a basis for diversity. *Plant Cell*, 14, s111-s130.
- Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics*, 98, 625-640.

- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nordborg M, Borevitz JO, Bergelson J, Berry C, Chory J *et al.* (13 co-auteurs) (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, 30, 190-193.
- O'Hara RB, Merilä J (2005) Bias and precision in *Qst* estimates: problems and some solutions. *Genetics*, 171, 1331-1339.
- Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52, 935-949.
- Orr HA (2005) The genetic theory of adaptation. *Nature Reviews Genetics*, 6, 119-127.
- Otto SP (2007) The evolutionary consequences of polyploidy. *Cell*, 131, 452-462.
- Paillard S (1999) Evolution des résistances à différentes maladies dans des populations de blé tendre (*Triticum aestivum* L.) menée en gestion dynamique. Thèse, Institut National Agronomique Paris-Grignon.
- Paillard S, Goldringer I, Enjalbert J, Doussinault G., de Vallavieille-Pope C, Brabant P (2000) Evolution of resistance against powdery mildew in winter wheat populations conducted under dynamic management. I- Is specific seedling resistance selected ? *Theoretical Applied Genetics*, 101,449-456.
- Paillard S, Goldringer I, Enjalbert J, Trottet M, David J, De Vallaveille-Pope C, Brabant P (2000b) Evolution of resistance against powdery mildew in winter wheat populations conducted under dynamic management. II- Adult plant resistance. *Theoretical and Applied Genetics*, 101, 457-462.
- Palo JU, O'Hara RB, Laugen AT, Laurila A, Primmer CR, Merilä J (2003) Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular Ecology*, 12, 1963-1978.
- Petit C, Thompson JD (1998) Phenotypic selection and population differentiation in relation to habitat heterogeneity in *Arrhenatherum elatius* (Poaceae). *Journal of Ecology*, 86, 829-840.
- Picard E., 1984. Contribution à l'étude de l'hérédité et de l'utilisation en sélection de l'haplodiploïdisation par androgénèse in vitro chez une céréale autogame *Triticum aestivum* L., Thèse d'Etat, Université Paris Sud.
- Pompanon F, Bonin A, Bellemain E, Taberlet P(2005) Genotyping errors: causes, consequences and solutions. *Nature Review Genetics*, 6, 847-859.
- Porcher E, Giraud T, Lavigne C (2006) Genetic differentiation of neutral markers and quantitative traits in predominantly selfing metapopulations: confronting theory and experiments with *Arabidopsis thaliana*. *Genetical Research*, 87, 1-12.

- Price AH (2006) Believe it or not, QTLs are accurate ! Trends in Plant Science, 11, 213-216.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.
- Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. Evolution, 59, 2312-2323.
- Putterill J, Laurie R, Macknight R (2004) It's time to flower: the genetic control of flowering time. BioEssays, 26, 363-373.
- Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology, 5, 94-100.
- Raquin AL (2005) Evolution de la diversité microsatellite dans une population de blé tendre conduite en gestion dynamique : effets de la mutation, de la dérive et de la sélection. Thèse, Université Paris XI, Orsay.
- Raquin AL, Depaulis F, Lambert A, Galic N, Brabant P, Goldringer I (in prep) Experimental estimation of mutation rates in a wheat population with coalescent-related methods.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.
- Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theoretical and Applied Genetics, 106, 895-902.
- Remington DL, Purugganan MD (2003) Candidate genes, Quantitative trait loci, and functional trait evolution in plants. International Journal of plant science, 164, s7-s20.
- Reynolds MP, Ortiz-Monasterio JI, and McNab A (2001) Application of Physiology in Wheat Breeding. Mexico, D.F.: CIMMYT.
- Rhoné B, Raquin, AL, Goldringer I (2007) Strong linkage disequilibrium near the selected *Yr17* resistance gene in a wheat experimental population. Theoretical and Applied Genetics, 114, 787-802.
- Rhoné B, Remoué C, Galic N, Goldringer I, Bonnin I (2008). Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. Molecular Ecology, 17, 930-943.
- Robert D, Gate P, Couvreur F (1993) Les satdes du blé. Institut technique des céréales et des fourrages, La Primaube.
- Röder M, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal M (1998) A microsatellite map of wheat. Genetics, 149, 2007-2023

- Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, 14, 351-361.
- Roux F, Touzet P, Cuguen J, Le Corre V (2006) How to be early flowering: an evolutionary perspective. *Trends in Plant Science*, 11, 375-381.
- Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S et al. (1995) Microsatellite evolution – Evidence for directionality and variation in rate between species. *Nature Genetics*, 10, 337-343.
- Sandring S, Riihimäki MA, Savolainen O, Ågren J (2007). Selection on flowering time and floral display in an alpine and a lowland population of *Arabidopsis lyrata*. *Journal of Evolutionary Biology*, 20, 558-567
- SAS. 2002. SAS version 8 for Windows. SAS Institute, Cary, North Carolina, USA.
- Sherman JD, Yan L, Talbert L, Dubcovsky J (2004) A PCR marker for growth habit in common wheat based on allelic variation at the *VRN-A1* gene. *Crop Science*, 44, 1832–1838.
- Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, 160, 753-763.
- Siol M, Bonnin I, Olivieri I, Prosperi JM, Ronfort J (2007) Effective populations size associated with self-fertilization: lessons from temporal changes in allele frequencies in the selfing annual *Medicago truncatula*. *Journal of Evolutionary Biology*, 20, 2349-2360.
- Slate J (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular Ecology*, 14, 363-379.
- Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theoretical Applied Genetics*, 109, 1105-1114.
- Song QJ, Shi JR, Singh S, Fickus EW, Costa JM et al. (2005) Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical Applied Genetics*, 110, 550-560.
- Sourdille P, Snape JW, Cadalen T, Charmet G, Nakata N, Bernard S, Bernard M (2000) Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. *Genome* 43:487-494
- Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, 135, 367-374.
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, 100, 158-170.
- Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology and Evolution*, 18, 94-101.

- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, 14, 671-688.
- Strasburg JL, Gross BL (2008) Adapting to winter in wheat: a long-term study follows parallel phenotypic and genetic changes in three experimental wheat populations. *Molecular Ecology*, 17, 216-218.
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceeding of the National Academy of Sciences*, 16, 9161-9166.
- Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps ? *Genome research*, 16, 702-712.
- Thomas G, Rousset M, Pichon M, Trottet M, Doussinault G, Picard E (1991) Méthodologie de l'amélioration de blé tender (*Triticum aestivum* L.) I. Création par croisements et analyse d'une population artificielle à 16 parents, base de cette étude méthodologique. *Agronomie*, 11, 359-368.
- Thuillet AC, Bru D, David JL, Roumet P, Santoni S et al (2002) Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. *Ssp durum desf.* *Molecular Biology and Evolution*, 19, 122-125.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC et al. (19 co-auteurs) (2006) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, 39, 31-40.
- Trottet M., 1988. Use of genic male sterility for breeding wheat lines resistant to *Leptosphaeria nodorum* Müller : results of a first selection cycle and prospects, In : Seventh International Wheat Genetics Symposium, Cambridge, UK, 13-19 July 1988, Institut of plant Science Research, 1199-1202.
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, 14, 3623-3642.
- Veyriéras JB (2006) Etude du déterminisme génétique de caractères quantitatifs chez les végétaux : méta-analyses de QTL et étude d'association. Thèse, Institut National Agronomique, Paris-Grignon.
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD (2002) Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, 19, 1251-1260.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, 158, 1811-1823.
- Waldmann P, Garcia-Gil MR, Sillanpää MJ (2005). Comparing Bayesian estimates of genetic differentiation of molecular markers and quantitative traits: an application to *Pinus sylvestris*. *Heredity*, 94, 623-629.

- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121, 379-391.
- Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracillaria gracilis* (Rhodophyta). *Molecular Ecology*, 7, 1569-1573.
- Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics*, 35, 235-254.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38, 1358-1370.
- Whitlock MC (1999). Neutral additive genetic variance in a metapopulation. *Genetical Research*, 74, 215-221.
- Worland AJ (1996) The influence of flowering time genes on environmental adaptability on European wheats. *Euphytica*, 89, 49-57.
- Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8, 206-216.
- Wright S (1931) Evolution in Mendelian Populations. *Genetics*, 16, 97-151.
- Wright S (1952) The theoretical variance within and among subdivisions of a population that is in a steady state. *Genetics*, 37, 312-21.
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. *Proceedings of the National Academy of Sciences*, 100, 6263-6268.
- Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J (2004) Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theoretical and Applied Genetics*, 109, 1677-1686.

# *ANNEXES*

# **Annexe 1 :**

## **Protocoles**

- A. Protocole d'extraction d'ADN
- B. Protocole de migration des produits PCR pour le génotypage des marqueurs microsatellite
- C. Protocole de PCR pour le génotypage des marqueurs microsatellite
- D. Conditions d'amplification des différents polymorphismes de gènes candidats

### Objectifs :

Extraire de l'ADN de haute qualité pour la PCR, à partir de poudre de feuilles lyophilisée (ou fraîche). Avec une plaque Whatman, on obtient de quoi faire une cinquantaine de réactions PCR.

### Principe Général de la manipulation.

- Extraction des composés hydrosolubles en présence de métabisulfite, puis précipitation d'une partie des protéines par l'acétate de potassium et centrifugation pour éliminer les débris insolubles.
- Fixation des acides nucléiques sur une membrane de silice en présence d'ions chaotropes (chlorure de guanidium), lavage de la membrane, puis élution de l'ADN.

### Domaine d'application :

- Toutes applications reposant sur la PCR, pour lesquelles la qualité de l'ADN est importante, mais il n'est pas nécessaire d'obtenir de grosses quantités (à exclure pour des blots).
- En particulier : amplification PCR de gènes pour séquençage ou génotypage SNP.

### Matériel :

- 20mg de feuilles broyées à froid puis lyophilisées aliquotées dans une plaque DeepWell 2mL.
- Pipette multicanaux P300 avec cônes adaptés. (**Ne pas utiliser les cônes habituels de P200**)
- Centrifugeuse Héraus à 4°C
- Bain marie à 65°C
- 2 plaques deep well 2mL
- 2 plaques PCR 96
- 2 plaques de filtration (membrane laine de verre) Whatman unifilter 800 GF/B (Whatman Ref 7700-2803)

### Réactifs :

#### 1) Tampon d'extraction :

Volume final	100 ml	200 ml	300 ml	400ml
Tris 1M pH = 8	10ml	20	30	40
EDTA 0,5M	10ml	20	30	40
NaCl 2.5M	20ml	40	60	80
SDS 20%	6,25ml	12,5	19	25
PVP 40000	1 g	2 g	3 g	4 g
Na Bisulfite	1 g	2 g	3 g	4 g
ddH <sub>2</sub> O	53.5ml	107ml	161ml	215

Ajouter le NaBisulfite et la RNase A (Amersham Ref XXXX) au moment de l'utilisation.

#### 2) Acétate de potassium: (3M K et 5M Ac)

Pour 100 ml : mélanger 60 ml de KAc 5M, 11.5 ml d'acide acétique glacial et 28.5 ml d'H<sub>2</sub>O.

#### 3) Tampon CGE (Chlorure de Guanidium 7.8M / Ethanol)

- Solution de Chlorure de guanidium 7.8 M . (Sigma Ref G-3272)

Dissoudre 37.25 gr de chlorure de guanidium dans l'eau jusqu'à un volume final de 50 ml de solution (faire le mélange dans un tube Sarstedt de 50 ml).

**- Le tampon CGE est obtenu en mélangeant 1/3 de solution de chlorure de guanidium 7.8 M et 2/3 d'éthanol à 96%.**

#### 4) Tampon de lavage (éthanol et sels d'acétate de potassium)

- Faire une solution aqueuse (Acétate de potassium 160 mM, Tris HCl ph 8 22,5 mM, EDTA 0,1 mM)

Volume final	100 ml	200 ml	300 ml	400ml
Tris 1M pH = 8	2,25 ml	4,5 ml	6,75 ml	9 ml
EDTA 0,5M	20 µl	40 µl	60 µl	80 µl
Acétate de potassium 5 m	3,2 ml	6,4 ml	9,6 ml	12,8 ml
ddH <sub>2</sub> O	94,50 ml	189 ml	283,50 ml	378 ml

- Le tampon de lavage est obtenu en mélangeant 100 ml de cette solution et 170 ml d'éthanol à 96%.

### Hygiène et sécurité.

- Le tampon CGE contient du chlorure de guanidium qui est irritant. Porter des gants et éviter tout contact avec la peau.

### Déroulement de la manipulation

**DUREE** : 3 Heures sans compter l'aliquotage des poudres (environ 1 heure 1/2 par plaque) et la préparation des solutions.

Personne de référence pour ce protocole :Xavier RAFFOUX.

#### 1) Phase d'extraction :

Préchauffer le tampon d'extraction à 65°C.

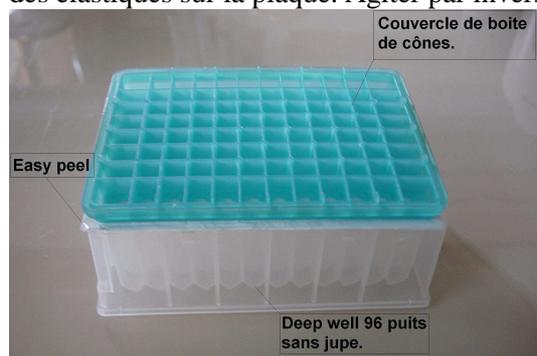
Ajouter le Na Bisulfite : 1g pour 100mL

Repartir 1000µL de tampon d'extraction dans chaque puit.

Fermer la plaque Deep Well avec un film EasyPeel (appuyer sur le film en frottant avec un kleenex).

Agiter doucement par inversion en maintenant le film EasyPeel avec un kleenex et un couvercle de boîte de cône.Le tout maintenu avec des élastiques.

Incuber à 65°C pendant 30 min au bain marie en plaçant un couvercle de boîte de cône maintenu par des élastiques sur la plaque. Agiter par inversion toutes les 10 minutes.



#### 2) Déprotéinisation et filtration

Ajouter 300 µl(avec multidispenser eppendorf seringue 5mL) d'acétate de K (5M/3M) froid (dans la glace).Sceller la plaque avec un film EasyPeel. Agiter doucement par inversion en utilisant un support de boîte de cône pour maintenir le film easy peel. Incuber 5 min dans la glace.

Centrifuger 10 min à 5600 rpm à 4°C, rotor 6444.

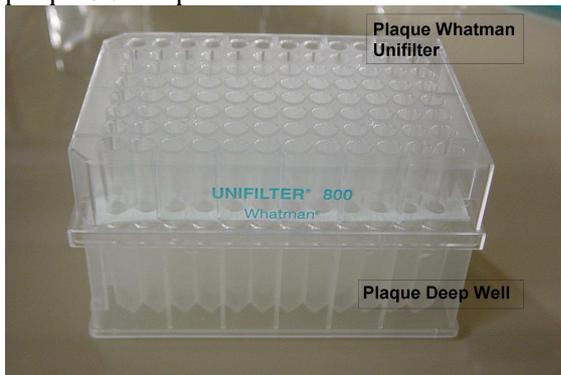
Transférer 800 µl (4X200µL avec multi eppendorf)de surnageant dans une plaque DeepWell 96 propre.

### 3) Fixation de l'ADN

Ajouter 1000 µl de tampon CGE (avec multidispenseur et seringue 10mL).

Mélanger soigneusement par pipetage.

Appliquer 800 µl du mélange sur chaque plaque Whatman unifilter (avec multi eppendorf). Centrifuger les plaques 2 min à 5600 rpm à T° ambiante. Eliminer le liquide qui passe dans une vieille plaque 96 Deepwell.

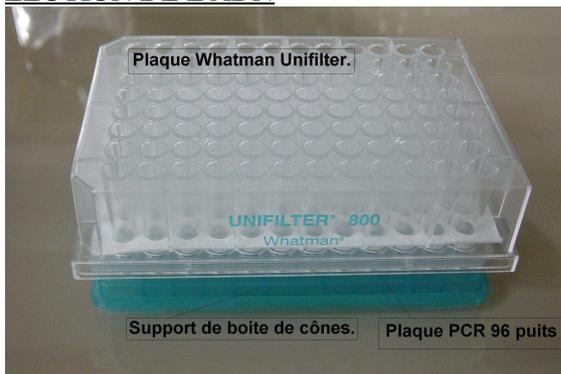


### 4) Lavage

1° lavage: déposer 500 µl de tampon de lavage (Ethanol et sel d'acétate de potassium) sur chaque colonne. Centrifuger les plaques 2 min à 5600 rpm à T° ambiante. **Eliminer le liquide qui passe.**

2° lavage: déposer 500 µl de tampon de lavage (Ethanol et sels d'acétate de potassium) sur la colonne. Centrifuger les plaques 15 min à 5600 rpm à T° ambiante (pour bien essorer la colonne). **Eliminer le liquide qui passe.**

### ELUTION DE L'ADN



Positionner les plaques PCR 96 sur un support de boîte de cône bleue. Poser les colonnes sur cet assemblage.

Déposer sur chaque colonne 50 µl d'eau UP préchauffée à 65 °C. Attendre 5 min à T° ambiante. Centrifuger 2 min à 1000 rpm à T° ambiante. Récupérer le liquide qui passe sur les plaques PCR 96 en polypropylène

Déposer de nouveau sur la colonne 50 µl d'eau UP préchauffé à 65 °C. Attendre 5 min à T° ambiante. Centrifuger 2 min à 1000 rpm à T° ambiante. Récupérer le liquide qui passe sur les plaques PCR 96 en polypropylène.

Congeler les ADN natifs.

Faire un gel de contrôle agarose standard 0.8%



Référence	MO-GEN-010	 <i>Original : Copie N°</i>	Date de Rédaction	28/11/05
Numéro de Version	2		Diffusion Privée Contrôlée	<input type="checkbox"/>
Page	1/6		Diffusion Privée Non Contrôlée	<input type="checkbox"/>
			Diffusion Publique Contrôlée	<input type="checkbox"/>
			Diffusion Publique Non Contrôlée	<input type="checkbox"/>

## Réalisation d'une plaque dépôt et lancement du séquenceur ABI 3100.

## 1- Objet / Domaine d'application

Ce protocole est utile pour la dilution des plaques PCR ainsi que pour la réalisation de la plaque de dépôt en vue de l'analyse sur le séquenceur capillaire.

## 2- Matériel et Réactifs

### Matériel

- ABI3100 GENETIC ANALYSER (Applera)
- GENESIS WORKSTATION 150 (Tecan)
- Heat Sealer (Abgene)
- Centrifugeuse BR 4i (Jouan)
- Tétrades MJ
- Septa 384 (Applera) réf : 4315934
- Plaques 384 puits PCR naturel (Abgene) réf : 016026 ou plaque de stockage 200µl, AB-1178 réf : 16572
- Plaques 96 puits PCR naturel (Abgene) réf : 016015
- Plaque 384 Deepwell 250 µl (Abgene) réf : 016111
- Film Thermosoudable Easy Peel (Abgene) réf : 016279
- Tubes à vis 2 ml
- Mono Pipette Eppendorf 10 µl
- Mono Pipette Eppendorf 200 µl
- Mono Pipette Eppendorf 5000 µl

### Réactifs

- Mix Dépôt : tubes à vis 2 ml contenant chacun 600 µl HI-DI formamide (Applera) + 2 µl ROX Size Standard ou 500 LIZ (Applera)
- 3730 POP7POLYMER réf : 4335615 ou 3100 POP4 Polymère réf : 4316355 (Applera)
- PCR marquées 6-FAM, HEX, NED, VIC, PET
- Eau milliQ

**DANGER CHIMIQUE.** Le **formamide** est dangereux s'il est absorbé par la peau ; il peut également provoquer une irritation des yeux, de la peau et de l'appareil respiratoire. Il risque d'endommager le système nerveux central et les systèmes de reproduction féminin et masculin, et pose un risque possible de malformation congénitale.

Porter des protections oculaires, des gants des vêtements appropriés et travailler sous hotte.

**DANGER CHIMIQUE.** Le **polymère POP** risque de provoquer l'irritation des yeux, de la peau et de l'appareil respiratoire. Porter des protections oculaires, des gants et des vêtements appropriés.

**Le liquide contenu dans le réservoir poubelle du séquenceur (contenant le polymère) doit être vidé dans le flacon prévu à cet usage et les plaques de dépôt après passage sur séquenceur (contenant la formamide) dans la poubelle prévue également à cet usage.**

<u>Rédacteur :</u> Séverine Rougeol	<u>Vérificateur :</u> Caroline Pont	<u>Approbateur :</u> Gilles Charmet
--	--	--

<b>Référence</b>	MO-GEN-010	 <small>Institut National de la Recherche Agronomique</small> <small>Original : Copie N°:</small>	<b>Page</b>	2/6
<b>Réalisation d'une plaque dépôt et lancement du séquenceur ABI 3100.</b>				

### Documents associés

Protocole de maintenance du séquenceur ABI 3100 (**MO-GENO-005**)  
 Protocole de maintenance du séquenceur ABI 3100 avec changement du capillaire (**MO-GENO-006**)  
 Protocole de maintenance du TECAN (**MO-GENO-004**)

## 3- Etapes à respecter

### 3.1 Préparation de l'échantillon avec le TECAN

- ✓ Allumer le Tecan
- ✓ Remplir le bidon d'eau milliQ
- ✓ Ouvrir le logiciel GEMINI
- ✓ Cliquer sur flush pour que l'instrument s'initialise
- ✓ Faire un flush de 50 et 25 ml (avec l'option deep)
- ✓ Faire quelques flushes de 7 et 3 (avec l'option shallow)
- ✓ Ouvrir le programme approprié aux PCR utilisées :
  - Avec des PCR en plaques 96, utiliser le programme « poolage\_dilution\_dans384\_new.gem » (progTECAN/dépôt/PCR96)
  - Avec des PCR en plaque 384, utiliser les programmes (progTECAN/dépôt/PCR384) :
    - Pour 1 plaque PCR : « dilution\_3PCR384dans384 » et « preparation\_depot\_384dans384 »
    - Pour 2 plaques PCR : « DepotRapide2PCRsering384 »
    - Pour 3 plaques PCR : « dvtnewpoolageSebSering384 »
    - Pour 4 plaques PCR : « poolage4PCR »
- ✓ Prendre une plaque 384 Deepwell 200 µl remplie avec de l'eau milliQ (stockée à 4°C), la poser en position (18/2)
- ✓ Centrifuger (faire un pulse) les plaques de PCR, enlever le film et les déposer sur le premier portoir comme indiqué dans le programme.
- ✓ Lancer le programme
- ✓ Sortir du congélateur 8 tubes de formamide. Une fois décongelés, ajouter 2 µl de marqueur de taille appropriés (350/400/500), vortexer et centrifuger avec la centrifugeuse de pailleasse.
- ✓ Placer les tubes sur le portoir comme indiqué dans le programme.
- ✓ Continuer le programme.

A la fin du programme, centrifuger (faire un pulse) la plaque de dépôt puis sceller un easy peel. Si la plaque n'est pas dénaturée tout de suite, la placer au congélateur.

### 3.2 Dénaturation des échantillons

Pour dénaturer les échantillons : lancer le programme « Denature ». Les échantillons sont chauffés à 95°C pendant 5 min puis les laisser à 4°C pendant au minimum 5 min.

### 3.3 Préparation d'un ensemble plaque

- Centrifuger (faire un pulse) la plaque de dépôt.
  - Placer un tapis de bouchons à septum propre et sec sur la plaque d'échantillons
- IMPORTANT** Le tapis de bouchons à septum doit être à plat sur la plaque.
- Placer la plaque d'échantillons sur le socle.
  - Fixer le clip de maintien à la plaque et au socle.

<b>Rédacteur :</b> Séverine Rougeol	<b>Vérificateur :</b> Caroline Pont	<b>Approbateur :</b> Gilles Charmet
--	--	--

Référence	MO-GEN-010	 Institut National de la Recherche Agronomique <i>Original : Copie N°:</i>	Page	3/6
			<b>Réalisation d'une plaque dépôt et lancement du séquenceur ABI 3100.</b>	

- Vérifier que les trous du clip de maintien et du tapis de bouchons à septum sont alignés.
- Centrifuger (faire un pulse) la plaque de dépôt sur son socle.

**IMPORTANT** Les pointes des barrettes risquent d'être endommagées si le clip de maintien et les trous du tapis de bouchons ne sont pas correctement alignés. Les trous du clip de maintien doivent s'aligner avec le tapis de bouchons à septum.

### 3.4 Vérification et remplissage des liquides

- **Si le polymère sur l'instrument** a moins d'une semaine et dispose d'une quantité insuffisante (<700µl) pour terminer les séries d'analyse, remplir les seringues et le bloc polymère supérieur de polymère conformément au protocole de maintenance.

**IMPORTANT** Toujours remplacer le polymère âgé de plus d'1 semaine, il est important de toujours gérer au mieux la quantité de POP dans les seringues. Lors du remplissage, le nombre de plaques doit être prévu.

**IMPORTANT** Vérifier l'absence de bulles d'air dans le bloc de polymère supérieur avant de poursuivre. Pour éliminer les bulles d'air, se reporter au protocole de maintenance.

Le polymère a un coût élevé, il ne doit pas être gaspillé, aussi bien pendant le remplissage des seringues que lorsque l'on chasse les bulles d'air.

- **Si le polymère sur l'instrument** a moins d'une semaine, et dispose d'une quantité suffisante pour terminer les séries d'analyse (Le passage d'une plaque 384 complète nécessite 700µl). Vérifier l'absence de bulles d'air, et procéder à la préparation de l'instrument.

- **Changement de tampon :**

Remplacer tous les 2 jours le tampon 1X dans les réservoirs de tampon anodique et de tampon cathodique.

**IMPORTANT** Le non remplacement du tampon risque de provoquer une perte de résolution et de la qualité des données.

**IMPORTANT** Pour le remplissage du tampon et le positionnement de la plaque, le passeur d'échantillons doit être placé en position avant, les pointes capillaires étant retirées de la solution tampon. Ne pas laisser le passeur d'échantillons dans cette position pendant plus **de 30 mn** car les capillaires peuvent sécher.

#### Préparation du tampon

- préparer 25 ml de tampon 1X
- Pour remplir les réservoirs d'eau et de tampon cathodique :
  - Refermer les portes de l'instrument.
  - Appuyer sur **Tray** sur le panneau externe de l'instrument pour amener le passeur d'échantillons en position avant.
  - Attendre l'arrêt du passeur d'échantillons, puis ouvrir les portes de l'instrument.
  - Retirer les réservoirs d'eau et de tampon cathodique de l'instrument.
  - Jeter les liquides restants et rincer les réservoirs à l'eau milliQ.

**Remarque** Il faut respecter les pratiques d'évacuation des déchets appropriées : **Le liquide contenu dans le réservoir poubelle du séquenceur (contenant le polymère) doit être vidé dans le flacon prévu à cet usage.**

- Rincer et sécher le réservoir cathodique, puis remplir jusqu'au repère avec le tampon 1X (environ 17 ml). Attention à ne pas mettre de gouttes sur les parois.
- Remplir les réservoirs d'eau jusqu'à la ligne avec de l'eau milliQ de qualité (environ 17 ml).
- Placer une bande de bouchons à septum propre sur chaque réservoir, et sécher la surface externe des réservoirs à l'aide d'un chiffon non pelucheux.

<b>Rédacteur :</b> Séverine Rougeol	<b>Vérificateur :</b> Caroline Pont	<b>Approbateur :</b> Gilles Charmet
--	--	--

Référence	MO-GEN-010	 Institut National de la Recherche Agronomique Original : Copie N°:	Page	4/6
			<b>Réalisation d'une plaque dépôt et lancement du séquenceur ABI 3100.</b>	

**Remarque** Nous conseillons d'étiqueter les réservoirs pour éviter de les mélanger. Le tapis de bouchons à septum doit être ajusté de façon serrée et à niveau avec le dessus des réservoirs pour ne pas endommager les pointes capillaires.

**ATTENTION !** La bande de bouchons doit reposer à plat sur le réservoir

- Remplissage du réservoir de tampon anodique

Changer le tampon anodique :

Avant chaque série, ou au moins toutes les 48 heures.

A chaque remplissage du bloc polymère avec du nouveau polymère

Placer les réservoirs en position sur le passeur d'échantillon comme indiqué ci-dessous.

*Pour remplir le réservoir de tampon anodique jusqu'au repère avec le tampon 1X :*

- Enlever le réservoir de tampon anodique en le poussant fermement vers le bas d'un lent mouvement de torsion.
  - Jeter le tampon.
  - Nettoyer et rincer et sécher le réservoir à l'eau milliQ.
  - Remplir le réservoir jusqu'au repère de remplissage avec un tampon 1X frais (environ 8 ml).
  - Placer le réservoir de tampon anodique sur l'instrument une fois que les bulles sont chassées.
- Remarque** Le ménisque doit s'aligner avec la ligne repère.

### 3.5 Positionnement de la plaque sur le passeur d'échantillons

- Positionner la plaque sur l'autosampler en faisant attention à l'encoche après avoir fait un **Tray**.
- Lorsque la plaque est correctement positionnée, l'indicateur de positionnement de plaque sur la page **Plate View** (Fenêtre de plaque) passe du gris au jaune. Vérifier que ce changement s'est bien produit.
- Refermer les portes de l'instrument.

**Remarque** Il n'y a qu'une orientation pour la plaque : le côté encoché du socle doit être vers le fond du passeur. **IMPORTANT** Vérifier que l'ensemble s'ajuste à plat sur le passeur d'échantillons.

Si cette mesure n'est pas observée, les pointes capillaires risquent de soulever les plaques du passeur d'échantillons.

**Remarque** La fermeture des portes renvoie le passeur d'échantillons dans la dernière position qu'il occupait avant l'ouverture des portes.

### 3.6 Création d'un enregistrement de plaque

**Remarque** On ne peut pas créer d'enregistrement de plaque tant qu'une analyse est en cours.

- Cliquer sur l'onglet **Plate View** (Fenêtre de plaque) de la fenêtre **3100 Data Collection Software** pour passer à la page **Plate View**.
- Importer la feuille de route en format texte.
- Ouvrir le tableur **Plate Editor** en faisant un double clique sur le nom de la feuille de route.
- Sélectionner un nom de projet sur la liste déroulante.
- Pour attribuer le même nom de projet à chaque échantillon de l'enregistrement de plaque. Cliquer sur l'en-tête de colonne pour sélectionner toute la colonne. Appuyer sur **CTRL+D**.

<b>Rédacteur :</b> Séverine Rougeol	<b>Vérificateur :</b> Caroline Pont	<b>Approbateur :</b> Gilles Charnet
--	--	--

## Réalisation d'une plaque dépôt et lancement du séquenceur ABI 3100.

- Pour chaque échantillon, sélectionner le **Run Module** (Module d'électrophorèse) approprié dans la liste déroulante.
- Pour chaque échantillon, sélectionner le **Analysis Module** (Module d'analyse) approprié dans la liste déroulante.
- Si on utilise l'étalon de masse moléculaire... Sélectionner ce module d'analyse...  
400HD:GS400HDgb.gsp / GS350:GS350gb.gsp / GS500:GS500gb.gsp
- Vérifier que l'enregistrement de plaque est correct, et cliquer sur **OK**.

Pour le logiciel d'analyse **ABI PRISM**® **GeneScan**®, sélectionner le jeu de fluorophores **Dye G5 et D**.  
**IMPORTANT** Vérifier que le jeu de fluorophores sélectionné convient pour la ou les séries d'analyse. Si les données sont recueillies en ayant sélectionné un mauvais jeu de fluorophores, les analyses doivent être recommencées.

Si l'opérateur souhaite analyser de nouveau le même échantillon, sélectionner un deuxième module d'électrophorèse et un deuxième module d'analyse. Un échantillon dans une plaque liée peut être analysé cinq fois au total. Pour cela, dans **Plate Manager**, sélectionner la feuille de route, faire **Edit**. Faire **Edit /Add Sample Run**, sélectionner dans les listes déroulantes **Result Group 2** et **Instrument Protocol 2**. Faire **Alt+D** pour relancer un run ou **Ctrl+D** pour une plaque complète.

**Remarque** Patienter pendant que le nouvel enregistrement de plaque est consigné dans la base de données et ajouté au tableau **Pending Plate Records** (Enregistrements des plaques provisoires) comme indiqué ci-dessous.

**Remarque** Pour qu'un enregistrement de plaque utilise un nom déjà utilisé, l'ancien enregistrement doit d'abord être supprimé de la base de données.

### 3.7 Liaison d'une plaque

**IMPORTANT** Une plaque peut être liée même si aucun module d'électrophorèse n'a été sélectionné pour ses échantillons. Dans ce cas, aucun message d'erreur n'apparaît et aucune série d'analyse ne sera programmée pour les échantillons de la plaque.

- Cliquer sur l'onglet **Plate View** (Fenêtre de plaque) de la fenêtre **3100 Data Collection Software** pour passer à la page **Plate View** et cliquer sur **Find All**
- Dans le tableau **Pending Plate Records**, cliquer sur la feuille de route associée à la plaque qui doit être liée.
- Et Cliquer sur le témoin de positionnement de plaque (en jaune) correspondant à la feuille de route.
- Vérifier que la plaque a bien été liée (la position de la plaque s'affiche devant le nom de la feuille de route).
- Le bouton **Run Instrument** (Lancer l'instrument) sur la barre d'outils est activé, ce qui signifie que l'instrument est prêt à fonctionner.
- Le témoin de positionnement pour la plaque liée devient vert.
- Répéter les étapes pour lier une deuxième plaque s'il y a lieu.
- L'enregistrement de plaque est dans le tableau des enregistrements de plaques liées.
- Cliquer sur l'onglet **Run View** (Fenêtre d'exécution) pour visualiser le calendrier de la série d'analyse (24 runs pour une plaque complète).

### 3.8 Démarrage et surveillance de la série

Cliquer sur le bouton vert **Run Instrument** (Lancer l'instrument) pour lancer les séries d'analyse programmées. Une série d'analyse utilisant **GeneScan\_POP4DefaultModule** prend environ 20 mn et 15 mn avec **GS22\_500\_POP7**.

- Cliquer sur l'onglet **Status View** (Fenêtre d'état) pour surveiller l'état de l'instrument pendant la série.

<b>Rédacteur :</b> Séverine Rougeol	<b>Vérificateur :</b> Caroline Pont	<b>Approbateur :</b> Gilles Charmet
--	--	--

Référence	MO-GEN-010	 <small>Institut National de la Recherche Agronomique</small> <small>Original : Copie N°:</small>	Page	6/6
<b>Réalisation d'une plaque dépôt et lancement du séquenceur ABI 3100.</b>				

- L'opérateur peut visualiser les données pendant l'analyse en utilisant les pages **Array Viewer** (Fenêtre de barrette) et **Capillary Viewer** (Fenêtre capillaire).

**IMPORTANT Toujours refermer les fenêtres Array Viewer et Capillary Viewer. Ne pas laisser ces fenêtres ouvertes pendant un temps prolongé lors d'une série ; des problèmes irrécupérables de mise à jour des fenêtres risquent de se produire.**

**Laisser la fenêtre 3100 Data Collection (les 2), orbix Web Daemon et AE Server ouverte.**

### 3.9 Arrêt d'une série et récupération des données

Quand une analyse est en cours, les boutons de passage à la série suivante, de pause et d'arrêt de série sont visibles sur la barre d'outils.

<b>Rédacteur :</b> Séverine Rougeol	<b>Vérificateur :</b> Caroline Pont	<b>Approbateur :</b> Gilles Charmet
--	--	--

Référence	MO-GEN-007	 <i>Original : ☒ Copie N°</i>	Date de Rédaction	15/05/2003
Numéro de Version	1		Diffusion Privée Contrôlée	<input type="checkbox"/>
Page	1/6		Diffusion Privée Non Contrôlée	<input type="checkbox"/>
			Diffusion Publique Contrôlée	<input type="checkbox"/>
			Diffusion Publique Non Contrôlée	<input type="checkbox"/>
<b>PCR 384 M13 en 6.5µl final pour génotypage SSR</b>				

## 1-Objet / Domaine d'application

Ce mode opératoire décrit les étapes de PCR en conditions M13 au format 384 puits en vue d'une analyse SSR sur séquenceur capillaire ABI 3100. Il doit être appliqué par tous les intervenants de la Plateforme Génotypage pour toute PCR M13 au format 384 puits.

## 2-Matériel et Réactifs

### Matériel

- × Plaques 384 puits PCR Abgene naturel non jupées (ADN)
- × Plaque DeepWell 96 800µl (mix).
- × Film Thermosoudable Easy Peel Abgene
- × Tubes à vis 2 ml (mix).
- × Mono Pipette Eppendorf 20 µl
- × Mono Pipette Eppendorf 200 µl
- × Mono Pipette Eppendorf 1000 µl
- × Heat Sealer Abgene
- × Centrifugeuse Jouan BR 4i
- × Robot pipeteur Biomek 2000 (mix).
- × Robot pipeteur Biomek FX (AND).
- × Tétrades MJ
- × Compress pad (facultative).

### Réactifs

- × Plaque(s) matrice(s) ADN (10 ng/ µl )
- × Taq DNA Polymérase Quiagen
- × Tampon 10X Quiagen
- × Mix dNTP Eurobio (mélange des 4 dNTPs à 5 mM chacun)
- × Primers (gauche avec extension M13, droite, M13 marquée)
- × H2O milliQ autoclavée

Rédacteur: Caroline PONT	Validateur : Gilles BOUTET	Approbateur : Gilles Charmet
-----------------------------	-------------------------------	---------------------------------

Référence	MO-GENO-007		
<b>PCR 384 en technique M13 pour génotypage SSR</b>			

Quantités pour une réaction (sans marge d'erreur) :

<i>[ ] Initiale</i>	<i>Volume</i>	<i>Quantités Finales</i>	<i>[ ] finales</i>
DNA (10 ng / $\mu$ l)	2.5 $\mu$ l	25 ng	3.8 ng / $\mu$ l
Taq DNA polymérase (5U/ $\mu$ l)	0,04 $\mu$ l	0,2U	0,031 U / $\mu$ l
Primer Mix (5 $\mu$ M droite, 0.5 $\mu$ M gauche M13)	0.655 $\mu$ l	3.27 pmoles droite 0.327 pmoles gaucheM13	50nM gM13 500 nM d
Primer M13 marquée (100 $\mu$ M)	0.034 $\mu$ l	3.4 pmoles	523 nM
dNTP mix 20 mM (5 mM of each)	0.27 $\mu$ l	5.4 nmoles (1.35 nmoles chaque)	830 $\mu$ M
MgCl <sub>2</sub> 25 mM (in tampon 10X)		15 nmoles	2.3 mM
Tampon 10X	0.65 $\mu$ l	6.5 X	1X
H <sub>2</sub> O	qsp 6.5 $\mu$ l (2.35 $\mu$ l)		

### 3-Etapes à respecter

#### Recommandations :

- Δ Mettre des gants et une blouse.
- Δ Travailler au maximum dans la glace et le plus rapidement possible.
- Δ Manipuler la fluorescence à l'abri de la lumière directe.

#### 3.1 Préparation du matériel.

- × Déstocker Tampon 10X, Mix dNTP, et primers. Les laisser décongeler dans la glace.
- × Préparer tout le matériel : étiqueter **tout** les tubes, plaques...
- × Préparer les amorces avec l'extension M13 :
  - (5  $\mu$ l amorce gauche M13 à 100  $\mu$ M+50  $\mu$ l amorce droite à 100  $\mu$ M+895  $\mu$ l H<sub>2</sub>O)
- × Préparer les plaques matrices ADN et réaliser un Quick-Run

Rédacteur: Caroline PONT	Validateur : Gilles BOUTET	Approbateur : Gilles Charmet
-----------------------------	-------------------------------	---------------------------------

Référence	MO-GENO-007	 Original :  Copie N°:	Page	-3/6-
<b>PCR 384 en technique M13 pour géotypage SSR</b>				

- \* Mettre en service les plateformes de pipetage Biomek 2000 et FX .
- \* Ouvrir les programmes appropriés sur chaque robot.
- \* Placer correctement les boîtes de Tips sur les plateformes.

### 3.2 Répartition du mix.

- \* Préparer un tube par marqueurs en respectant les quantités précisées en annexe 1.
- \* Vérifier l'emplacement de la boîte de tips et si le robot à été utilisé avant, s'assurer qu'il reste des tips, faire repartir la boite à zéro si nécessaire (option / reset all tips).
- \* Placer correctement la plaque PCR 384 sur le robot.
- \* Répartir le (ou les) premier(s) tube(s) de mix dans la (ou les) première(s) colonne(s) libre(s) de la Deep Well 800µl de sorte à réaliser une PCR entièrement. Faire un Quick-Run sur la DeepWell.
- \* Indiquer dans le logiciel avec quelle(s) colonne(s) le robot doit travailler.
- \* Lancer le robot en faisant start.

### 3.3 Transfert d' ADN dans les plaques PCR 384

- \* Pendant que le biomek 2000 répartit l'ADN, vérifier que tout est prêt sur le FX :
- \* Plaque d'ADN dans l'ordre avec A1 en haut à gauche.
- \* Idem pour les boîtes de tips.
- \* Mettre en route la station de lavage : baisser les 2 leviers, remplir le seau d'eau et vider la poubelle.
- \* Une fois la répartition du mix terminé faire un Quick-Run sur la plaque et la placer sur le FX.
- \* Lancer le programme en faisant start.
- \* Une fois la répartition de l'ADN terminé faire un Quick-Run sur la plaque et la sceller correctement (surtout sur les bords) avec un film Easy Peel .
- \* Inscrire sur le dessus le nom de la plaque de façon à ce que l'on sache à qui appartient cette plaque et à quand a t-elle été faite.
- \* Cycler cette plaque directement ou la conserver à -20°C.

<u>Rédacteur:</u> Caroline PONT	<u>Valideur :</u> Gilles BOUTET	<u>Approbateur :</u> Gilles Charmet
------------------------------------	------------------------------------	--

### 3.4 Cycling

- × Déstocker les plaques à cycler en fonction du « planning cycleur 384 »
- × Mettre en marche le cycleur
- × Réaliser un Quick-Run
- × Si possible préchauffé le bloc PCR à 95°C.
- × Mettre les plaques sur le cycleur avec un « compress pad » (facultatif).
- × Lancer le programme « M13-Tm » .

Initial Dénaturation :	95°C	5 min	
Dénaturation :	95°C	30 s	30 Cycles
Annealing :	Tm °C	30 s	
Extension :	72°C	30s	
Dénaturation :	95°C	30 s	8 Cycles
Annealing :	Tm -4°C	30 s	
Extension :	72°C	30s	
Final Extension:	72°C	5 min	
Storage :	4°C	inf.	

- × Remplir le cahier de cycling et signer - **INDISPENSABLE !!!**
- × Après cycling, inscrire un point sur le film de la PCR pour indiquer qu'elle a été cyclée.
- × Stocker la PCR à - 20 °C.

Rédacteur: Caroline PONT	Valideur : Gilles BOUTET	Approbateur : Gilles Charmet
-----------------------------	-----------------------------	---------------------------------

Référence	MO-GENO-007	Page	-5/6-
<b>PCR 384 en technique M13 pour génotypage SSR</b>			

**ANNEXE 1 : PREPARATION DU MIX PCR.**

**MIX POUR UN MARQUEUR PAR PCR 384 :**

	par Individ	1 plaque 384		2 plaques		3 plaques		4 plaques	
		1 marqueur		2 marqueurs		3 marqueurs		4 marqueurs	
		450		900		1350		1800	
<b>H2O</b>	2.35	<b>1058</b>		<b>2115</b>		<b>3173</b>		<b>4230</b>	
<b>Tp</b>	0.65	<b>293</b>		<b>585</b>		<b>878</b>		<b>1170</b>	
<b>dNTP (5µM)</b>	0.27	<b>122</b>		<b>243</b>		<b>365</b>		<b>486</b>	
total	3.27	1471.5	-4µl	2943	(.../2)-4µl	4414.5	(.../3)-4µl	5886	(.../4)-4µl
volume pipetté	3.27	<b>1471</b>	<b>x1</b>	<b>1467</b>	<b>x2</b>	<b>1467</b>	<b>x3</b>	<b>1467</b>	<b>x4</b>
<b>M13</b>	0.034	<b>15.3</b>		<b>15.3</b>		<b>15.3</b>		<b>15.3</b>	
<b>Taq</b>	0.04	<b>18</b>		<b>18</b>		<b>18</b>		<b>18</b>	
<b>primer (mix 0.5/5µM)</b>	0.655	<b>295</b>		<b>295</b>		<b>295</b>		<b>295</b>	
total	4	1800	(.../8) -10µl	1795	(.../8) -10µl	1795	(.../8) -10µl	1795	(.../8) -10µl
vol pipetté		<b>213µl</b>	<b>x8</b>	<b>213µl</b>	<b>x8</b>	<b>213µl</b>	<b>x8</b>	<b>213µl</b>	<b>x8</b>
vol réel besoin		192		192		192		192	

	par Individ	6 plaques		8 plaques		12 plaques	
		6 marqueurs		8 marqueurs		12 marqueurs	
		2700		3600		5400	
<b>H2O</b>	2.35	<b>6345</b>		<b>8460</b>		<b>12690</b>	
<b>Tp</b>	0.65	<b>1755</b>		<b>2340</b>		<b>3510</b>	
<b>dNTP (5µM)</b>	0.27	<b>729</b>		<b>972</b>		<b>1458</b>	
total	3.27	8829	(.../6)-4µl	11772	(.../8)-4µl	17658	(.../12)-4µl
volume pipetté	3.27	<b>1467</b>	<b>x6</b>	<b>1467</b>	<b>x8</b>	<b>1467</b>	<b>x12</b>
<b>M13</b>	0.034	<b>15.3</b>		<b>15.3</b>		<b>15.3</b>	
<b>Taq</b>	0.04	<b>18</b>		<b>18</b>		<b>18</b>	
<b>primer (mix 0.5/5µM)</b>	0.655	<b>295</b>		<b>295</b>		<b>295</b>	
total	4	1795	(.../8) -10µl	1795	(.../8) -10µl	1795	(.../8) -10µl
vol pipetté		<b>213µl</b>	<b>x8</b>	<b>213µl</b>	<b>x8</b>	<b>213µl</b>	<b>x8</b>
vol réel besoin		192		192		192	

<u>Rédacteur:</u> Caroline PONT	<u>Valideur :</u> Gilles BOUTET	<u>Approbateur :</u> Gilles Charmet
------------------------------------	------------------------------------	--

Référence	MO-GENO-007
-----------	-------------

**PCR 384 en technique M13 pour génotypage SSR**

**MIX POUR DEUX MARQUEURS PAR PCR 384 :**

	par Indiv	1 plaque 384		2 plaques		3 plaques		4 plaques	
		2 marqueurs		4 marqueurs		6 marqueurs		8 marqueurs	
		470		940		1410		1880	
<b>H2O</b>	2.35	<b>1105</b>		<b>2209</b>		<b>3314</b>		<b>4418</b>	
<b>Tp</b>	0.65	<b>306</b>		<b>611</b>		<b>917</b>		<b>1222</b>	
<b>dNTP (5µM)</b>	0.27	<b>127</b>		<b>253.8</b>		<b>381</b>		<b>507.6</b>	
<i>total</i>	3.27	1536.9	-4µl	3073.8	(.../2)-4µl	4610.7	(.../3)-4µl	6147.6	(.../4)-4µl
<b>volume pipetté</b>	<b>3.27</b>	<b>764</b>	<b>x2</b>	<b>764</b>	<b>x4</b>	<b>764</b>	<b>x6</b>	<b>764</b>	<b>x8</b>
<b>M13</b>	0.034	<b>8</b>		<b>8</b>		<b>8</b>		<b>8</b>	
<b>Taq</b>	0.04	<b>9.4</b>		<b>9.4</b>		<b>9.4</b>		<b>9.4</b>	
<b>primer (mix 0.5/5µM)</b>	0.655	<b>154</b>		<b>154</b>		<b>154</b>		<b>154</b>	
<i>total</i>	4	935	(.../8) -12µl	935	(.../8) -12µl	935	(.../8) -12µl	935	(.../8) -12µl
<b>vol pipetté</b>		<b>105</b>	<b>x8</b>	<b>105</b>	<b>x8</b>	<b>105</b>	<b>x8</b>	<b>105</b>	<b>x8</b>
<i>vol réel besoin</i>		96		96		96		96	

	par Indiv	6 plaques		8 plaques		12 plaques	
		12 marqueurs		16 marqueurs		24 marqueurs	
		2820		3760		5640	
<b>H2O</b>	2.35	<b>6627</b>		<b>8836</b>		<b>13254</b>	
<b>Tp</b>	0.65	<b>1833</b>		<b>2444</b>		<b>3666</b>	
<b>dNTP (5µM)</b>	0.27	<b>761.4</b>		<b>1015.2</b>		<b>1522.8</b>	
<i>total</i>	3.27	9221.4	(.../6)-4µl	12295	(.../8)-4µl	18443	(.../12)-4µl
<b>volume pipetté</b>	<b>3.27</b>	<b>764</b>	<b>x12</b>	<b>764</b>	<b>x16</b>	<b>764</b>	<b>x24</b>
<b>M13</b>	0.034	<b>8</b>		<b>8</b>		<b>8</b>	
<b>Taq</b>	0.04	<b>9.4</b>		<b>9.4</b>		<b>9.4</b>	
<b>primer (mix 0.5/5µM)</b>	0.655	<b>154</b>		<b>154</b>		<b>154</b>	
<i>total</i>	4	935	(.../8) -12µl	935	(.../8) -12µl	935	(.../8) -12µl
<b>vol pipetté</b>		<b>105</b>	<b>x8</b>	<b>105</b>	<b>x8</b>	<b>105</b>	<b>x8</b>
<i>vol réel besoin</i>		96		96		96	

<u>Rédacteur :</u> Caroline PONT	<u>Valideur :</u> Gilles BOUTET	<u>Approbateur :</u> Gilles Charmet
-------------------------------------	------------------------------------	--

# D

## Conditions d'amplification des différents polymorphismes de gènes

### I. Conditions d'amplification du promoteur de *VRN1A*

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,2 mM
Amorce F-M13	0,1 µM
Amorce R	0,1 µM
Queue M13	0,025 µM
Taq	1 U
ADN	25 ng
<b>Volume total</b>	<b>20 µl</b>

#### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	95°C	5 min	1
Dénaturation	95°C	20 sec	30
Hybridation	55°C	20 sec	
Elongation	72°C	30 sec	
Elongation finale	72°C	3 min	1
Conservation	4°C	10 min	
	10°C	infini	

Amorce F-**M13** : **CACGACGTTGTAAAACGAC**GAAAGGAAAAATTCTGCTCG

Amorce R : TGCACCTTCCSCGCCCAT

Taille attendue    Allèle 1    705 et 614 pb  
                           Allèle 2    483 pb  
                           Allèle 5    463 pb

### II. Conditions d'amplification de l'intron 1 de *VRN1*

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce R	0,5 µM
Taq	1 U
ADN	25 ng
<b>Volume total</b>	<b>22 µl</b>

## 2) Programme d'amplification

Pour le gène *VRN1B*

Amorce F : CAAGTGG AACGGTTAGGACA  
 Amorce R : CAAATGAAAAGGAATGAGAGCA  
 T° d'hybridation (Tm) : 56,4°C  
 Taille attendue : 1149 pb

*Non délétion*

*Délétion*  
 CAAGTGG AACGGTTAGGACA  
 CTCATGCCAAAAATGAAGATGA  
 58°C  
 709 pb

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	5 min	1
Dénaturation	94°C	30 sec	10
Hybridation	Incrément de 1°C à chaque cycle de Tm+10°C à Tm	30 sec	
Elongation	72°C	1 min 30	
Dénaturation	94°C	30 sec	28
Hybridation	Tm	30 sec	
Elongation	72°C	1 min 30	
Elongation finale	72°C	10 min	1
Conservation	4°C	10 min	
	10°C	infini	

Pour le gène *VRN1D*

Amorce F : GTTGTCTGCCTCATCAAATCC  
 Amorce R : AAATGAAAAGGAACGAGAGCG  
 Taille attendue : 1671 pb

*Non délétion*

*Délétion*  
 GTTGTCTGCCTCATCAAATCC  
 GGTCAGGTGGTCTGTGC  
 997 pb

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	5 min	1
Dénaturation	94°C	30 sec	10
Hybridation	Incrément de 1°C à chaque cycle de 71°C à 61°C	30 sec	
Elongation	72°C	1 min 30	
Dénaturation	94°C	30 sec	28
Hybridation	61°C	30 sec	
Elongation	72°C	1 min 30	
Elongation finale	72°C	10 min	1
Conservation	4°C	10 min	
	10°C	infini	

### III. Conditions d'amplification de l'intron 8 de *VRN1A*

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce R	0,5 µM
Taq	1 U
ADN	25 ng
Volume total	22 µl

#### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	4 min	1
Dénaturation	94°C	1 min	10
Hybridation	Incrément de 1°C à chaque cycle de 65°C à 55°C	1 min	
Elongation	72°C	1 min 30	
Dénaturation	94°C	1 min	21
Hybridation	55°C	1 min	
Elongation	72°C	1 min 30	
Elongation finale	72°C	5 min	1
Conservation	4°C	5 min	
	10°C	infini	

Amorce F : TCAGATTCTAGACTGAGATGTTCAA  
 Amorce R : GATGTGGCTCACCATCCACG  
 Taille attendue : 807 pb

Ensuite digestion du produit PCR par l'enzyme Aci I

### IV. Conditions d'amplification de l'intron 11 de *Luminidependens* génome B

#### IV.1 SNP1

##### IV.1.1 PCR génome spécifique

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce R	0,5 µM
Taq	1 U
ADN	25 ng
Volume total	11 µl

#### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	5 min	1
Dénaturation	94°C	30 sec	10
Hybridation	Incrément de 1°C à chaque cycle de 70°C à 60°C	30 sec	
Elongation	72°C	2 min	
Dénaturation	94°C	30 sec	20
Hybridation	60°C	30 sec	
Elongation	72°C	2 min	
Elongation finale	72°C	10 min	1
Conservation	4°C	10 min	
	10°C	infini	

Amorce F : GTGTCAGGTAATATGTATTGCTGC  
 Amorce R : ACCCAAGCCCATCAACTTCTAT  
 Taille attendue : 1474 pb

## IV.1.2 PCR allèle compétitive

### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F-M13	0,5 µM
Amorce F-M13-EcoRV	0,5 µM
Amorce R	0,5 µM
Taq	1 U
PCR 1 diluée 1/50	2 µl
Volume total	12 µl

### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	5 min	1
Dénaturation	94°C	30 sec	10
Hybridation	Incrément de 1°C à chaque cycle de 70°C à 60°C	30 sec	
Elongation	72°C	1 min	
Dénaturation	94°C	30 sec	20
Hybridation	60°C	30 sec	
Elongation	72°C	1 min	
Elongation finale	72°C	10 min	1
Conservation	4°C	10 min	
	10°C	infini	

Amorce F-M13 : CACGACGTTGTAACGACTTGTCATCTTGTGTACGAATTT  
 Amorce F-M13-EcoRV CACGACGTTGTAAGATATCTTGTATCTTGTGTACGAATTC  
 Amorce R : CTGGAAGAACACAAATAATGACG  
 Taille attendue : 144 pb

Ensuite digestion du produit PCR par l'enzyme EcoRV

## IV.2 SNP2

### IV.2.1 PCR génome spécifique

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce R	0,5 µM
Taq	1 U
ADN	25 ng
Volume total	22 µl

#### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	5 min	1
Dénaturation	94°C	30 sec	10
Hybridation	Incrément de 1°C à chaque cycle de 70°C à 60°C	30 sec	
Elongation	72°C	2 min	
Dénaturation	94°C	30 sec	20
Hybridation	60°C	30 sec	
Elongation	72°C	2 min	
Elongation finale	72°C	10 min	1
Conservation	4°C	10 min	
	10°C	infini	

Amorce F : GTGTCAGGTAATATGTATTGCTGC  
 Amorce R : ACCCAAGCCCATCAACTTCTAT  
 Taille attendue : 1474 pb

## IV.2.2 PCR allèle compétitive

### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce F-M13	0,25 µM
Amorce R	0,5 µM
Taq	1 U
PCR 1 diluée 1/20	2 µl
Volume total	22 µl

### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	5 min	1
Dénaturation	94°C	30 sec	10
Hybridation	Incrément de 1°C à chaque cycle de 71,5°C à 61,5°C	30 sec	
Elongation	72°C	1 min	
Dénaturation	94°C	30 sec	20
Hybridation	61,5°C	30 sec	
Elongation	72°C	1 min	
Elongation finale	72°C	10 min	1
Conservation	4°C	10 min	
		10°C	infini

Amorce F : TTATTCATAGCGTTGAAAATGCC  
 Amorce F-M13 : CACGACGTTGTAAAACGACTTATTCATAGCGTTGAAAATGCA  
 Amorce R : GCTTATATCAGGTTTGATCTCC  
 Taille attendue : 166 ou 185 pb

## V. Conditions d'amplification de *Gigantea* génome B

### V.1 PCR génome spécifique

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
Tampon Qiagen	1X
MgCl <sub>2</sub>	1,5 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce R	0,5 µM
Taq	0,5 U
ADN	25 ng
Volume total	25 µl

#### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	4 min	1
Dénaturation	94°C	1 min	10
Hybridation	Incrément de 1°C à chaque cycle de 66°C à 56°C	1 min	
Elongation	72°C	2 min 30	
Dénaturation	94°C	1 min	21
Hybridation	56°C	1 min	
Elongation	72°C	2 min 30	
Elongation finale	72°C	4 min	1
Conservation	10°C	infini	

Amorce F : CAGCCTCACACATGAAAAATG  
 Amorce R : ATTAGAATGTCACTGAAGCATTGA  
 Taille attendue : 372 pb

## V.2 PCR allèle compétitive

### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	3 mM
dNTP	0,6 mM
Amorce F ou F-M13-EcoRV	0,5 µM ou 0,25 µM
Amorce F-M13	0,25 µM
Amorce R	0,5 µM
Taq	1 U
PCR 1 diluée 1/50	2 µl
Volume total	25 µl

### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	4 min	1
Dénaturation	94°C	30 sec	30
Hybridation	60°C	30 sec	
Elongation	72°C	30 sec	
Elongation finale	72°C	5 min	1
Conservation	4°C	5 min	
	10°C	infini	

Amorce F : CAGAAGAGGCTAGGCGAAAC  
 Ou Amorce F-M13-EcoRV CACGACGTTGTAAGATATCCAGAAGAGGCTAGGCGAAAC  
 Amorce F-M13 : CACGACGTTGTAACGACCAGAAGAGGCTAGGCGAAAT  
 Amorce R : AGCCTCGATAACCCCAT

Lorsque l'amorce F-M13-EcoRV a été utilisée, digestion du produit PCR par EcoRV

## VI. Conditions d'amplification de *Constans* génome B

### VI.1 PCR génome spécifique

#### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
Tampon Qiagen	1X
MgCl <sub>2</sub>	3 mM
dNTP	0,1 mM
Amorce F	0,5 µM
Amorce R	0,5 µM
Taq	0,5 U
ADN	25 ng
Volume total	25 µl

#### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	4 min	1
Dénaturation	94°C	1 min	10
Hybridation	Incrément de 1°C à chaque cycle de 66°C à 56°C	1 min	
Elongation	72°C	2 min 30	
Dénaturation	94°C	1 min	21
Hybridation	56°C	1 min	
Elongation	72°C	2 min 30	
Elongation finale	72°C	4 min	1
Conservation	10°C	infini	

Amorce F : GTATGATGCACAAAATTCTAACCA  
 Amorce R : TGAGACAAGTACGATAAGATGGG  
 Taille attendue : 259 pb

## VI.2 PCR allèle compétitive

### 1) Mélange réactionnel PCR

Produits	Concentration/ Quantité finale
Tampon	1X
MgCl <sub>2</sub>	3 mM
dNTP	0,6 mM
Amorce F ou F-M13-EcoRV	0,5 µM ou 0,25 µM
Amorce F-M13	0,25 µM
Amorce R	0,5 µM
Taq	1 U
PCR 1 diluée 1/50	2 µl
Volume total	25 µl

### 2) Programme d'amplification

Etapes	Température	Temps	Nombre de cycles
Dénaturation	94°C	4 min	1
Dénaturation	94°C	30 sec	30
Hybridation	65°C	30 sec	
Elongation	72°C	30 sec	
Elongation finale	72°C	5 min	1
Conservation	4°C	5 min	
	10°C	infini	

Amorce F : CAAGCACTCTTGGTCGATCTC  
 Ou Amorce F-M13-EcoRI CACGACGTTGTAAGAATTCCAAGCACTCTTGGTCGATCTC  
 Amorce F-M13 : CACGACGTTGTAACGACCAAGCACTCTTGGTCGATCTT  
 Amorce R : AATTCTATGGACTTGATCTCTAT

Lorsque l'amorce F-M13-EcoRI a été utilisée, digestion du produit PCR par EcoRI

## **Annexe 2**

Annexe 2 : Ensemble des haplotypes multilocus reconstruits à partir des 7 locus microsattélites du chromosome 3B et répartition dans les populations étudiées.

Par	G0	Le Moulon			Toulouse			Vervins		
		G2	G7	G12	G2	G7	G12	G2	G7	G12
h1	2									
h2	2	1	2		2					
h3	2	1		2		2	2			
h4	2									
h5	2									
h6	2	1			2			2		
h7	2									
h8	2				2	4	1		2	
h9	2						1			
h10	2	2	2		2	10	22	5	2	
h11	4		2	2						
h12	2	2		2				2		
h13	2			2						
h14	2				2		2			
h15	3			4		17	4	6	2	
h16	4			4						
h17	2		4	2						
h18	6			4		2			2	
h19	2	2						4		
h20	1									
h21	2		2							
h22	2				3	12	2			
h23	2									
h24	4	2	4		6		2			
h25	2									
h26	2			1						
h27	2									
h28	3									
h29	1									
h30	1									
h31	2									
h32	2	2	2		4			2		

Par	G0	Le Moulon			Toulouse			Vervins		
		G2	G7	G12	G2	G7	G12	G2	G7	G12
h33	2						1			
h34	2						4			
h35	2								1	
h36	2									1
h37	2									
h38	2									
h39	2									
h40	2									
h41	1									
h42	1									
h43	2									
h44	3						2	2		
h45	1									
h46	2									
h47	2									
h48	2									
h49	2									
h50	1									
h51	1									
h52	2									
h53	2									
h54	1									
h55	1									
h56	2									
h57	2									
h58	2									
h59	2									
h60	2									
h61	2			5						
h62	2									
h63	2			1	2	2	2	2		
h64	2			1	5	1	1	2		
h65	4			4						
h66	1			1						
h67	1			2						
h68	2			2						
h69	3			2						
h70	2				6	2	4		2	
h71	2				2					
h72	2				2					
h73	1									
h74	2									
h75	4									
h76	1									
h77	1									
h78	2								4	
h79	2									
h80	2									
h81	4			2						
h82	2									
h83	1									
h84	1									
h85	4				2		6	10	26	
h86	2									
h87	2									
h88	2						1			
h89	2									
h90	2			2						
h91	2			6					2	
h92	2									
h93	1				2					
h94	1			2			4			
h95	2									
h96	2									
h97	2								2	
h98	3									
h99	14									
h100	2					2				
h101	2									
h102	1									
h103	5				2	2				
h104	2				2					
h105	2									
h106	2					2				
h107	1				2					
h108	2									
h109	2									
h110	2									
h111	2									
h112	2									
h113	1									
h114	1									
h115	1									
h116	1									
h117	1							1		
h118	2									
h119	2									
h120	2									
h121	2				2	2				
h122	2									
h123	2									

Par	G0	Le Moulon			Toulouse			Vervins		
		G2	G7	G12	G2	G7	G12	G2	G7	G12
h124	1									
h125	1									
h126	2									
h127	2									
h128	2									
h129	2									
h130	2									
h131	2									
h132	2									
h133	2									
h134	2									
h135	2				2					
h136	2							4	8	
h137	2									
h138	2						2			
h139	2					3				
h140	2									
h141	2									
h142	6					2				
h143	2									
h144	2									
h145	2									
h146	2									
h147	2									
h148	2									
h149	2									
h150	2									
h151	2									
h152	1									
h153	1							2		
h154	2					2				
h155	2									
h156	2									
h157	2									
h158	2									
h159	2									
h160	2									
h161	2									
h162	2									
h163	2									
h164	2									
h165	2									
h166	2									
h167	1									
h168	1									
h169	2					2				
h170	2									
h171	2									
h172	2									
h173	2									
h174	1									
h175	2									
h176	2							1	1	
h177	2									
h178	2									
h179	2									
h180	2									
h181	2									
h182	2									
h183	2									

Par	G0	Le Moulon			Toulouse			Vervins		
		G2	G7	G12	G2	G7	G12	G2	G7	G12
h184	1									1
h185	1									1
h186	4									4
h187	2									2
h188	2									2
h189	1								1	
h190	1									1
h191	1									1
h192	2									2
h193	2									2
h194	1								1	
h195	1									1
h196	2									2
h197	2									2
h198	4									4
h199	1									1
h200	2									2
h201	2									2
h202	1									1
h203	2							2		2

## **Annexe 3**

**Annexe 3 : Moyenne, écart-type (SD), effectifs (N), valeurs minimum et maximum de chaque caractère dans chaque population.**

**Rappel des abréviations :** Hd=date d'épiaison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante, NG\_E=nombre de grains par épi, P1000G=poids de 1000 grains.

Caractère	Population	Pop. Parentale	Pop. Initiale	Le Moulon			Toulouse			Vervins		
	Génération		G0	G2	G7	G12	G2	G7	G12	G2	G7	G12
<b>HdNV</b>	<b>Moyenne</b>	<b>1562</b>	<b>1649</b>	<b>1682</b>	<b>1728</b>	<b>1820</b>	<b>1622</b>	<b>1302</b>	<b>1373</b>	<b>1839</b>	<b>1662</b>	<b>1656</b>
	<i>SD</i>	<i>468</i>	<i>431</i>	<i>358</i>	<i>395</i>	<i>314</i>	<i>440</i>	<i>415</i>	<i>437</i>	<i>281</i>	<i>370</i>	<i>338</i>
	N	15	34	48	52	48	47	54	48	54	53	57
	Min	920	868	887	891	976	919	864	879	995	947	1153
	Max	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000
<b>Hd4S</b>	<b>Moyenne</b>	<b>1132</b>	<b>1184</b>	<b>1208</b>	<b>1177</b>	<b>1260</b>	<b>1216</b>	<b>1068</b>	<b>1082</b>	<b>1262</b>	<b>1277</b>	<b>1305</b>
	<i>SD</i>	<i>235</i>	<i>183</i>	<i>196</i>	<i>158</i>	<i>198</i>	<i>223</i>	<i>127</i>	<i>144</i>	<i>195</i>	<i>240</i>	<i>233</i>
	N	16	35	53	55	53	50	55	53	55	55	58
	Min	894	908	955	962	996	878	913	893	975	952	1045
	Max	1854	1940	2000	1725	1825	2000	1560	1544	2000	2000	1909
<b>Hd8S</b>	<b>Moyenne</b>	<b>1071</b>	<b>1075</b>	<b>1104</b>	<b>1099</b>	<b>1096</b>	<b>1097</b>	<b>1062</b>	<b>1051</b>	<b>1135</b>	<b>1146</b>	<b>1163</b>
	<i>SD</i>	<i>66</i>	<i>54</i>	<i>55</i>	<i>50</i>	<i>51</i>	<i>77</i>	<i>48</i>	<i>44</i>	<i>62</i>	<i>67</i>	<i>56</i>
	N	16	35	54	55	54	51	55	53	55	55	58
	Min	985	972	986	955	980	946	953	973	995	981	1028
	Max	1208	1178	1268	1193	1215	1421	1170	1136	1291	1299	1320
<b>HdJC</b>	<b>Moyenne</b>	<b>2335</b>	<b>2395</b>	<b>2615</b>	<b>2595</b>	<b>2586</b>	<b>2426</b>	<b>2395</b>	<b>2184</b>	<b>2393</b>	<b>2578</b>	<b>2638</b>
	<i>SD</i>	<i>514</i>	<i>594</i>	<i>467</i>	<i>433</i>	<i>408</i>	<i>653</i>	<i>582</i>	<i>589</i>	<i>535</i>	<i>506</i>	<i>411</i>
	N	16	35	49	49	49	49	49	50	49	49	49
	Min	1537	1337	1452	1507	1753	1212	1241	1247	1357	1234	1457
	Max	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
<b>HdCHP</b>	<b>Moyenne</b>	<b>1211</b>	<b>1204</b>	<b>1239</b>	<b>1240</b>	<b>1233</b>	<b>1219</b>	<b>1208</b>	<b>1197</b>	<b>1257</b>	<b>1271</b>	<b>1284</b>
	<i>SD</i>	<i>60</i>	<i>54</i>	<i>44</i>	<i>41</i>	<i>54</i>	<i>58</i>	<i>45</i>	<i>50</i>	<i>56</i>	<i>64</i>	<i>51</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	1118	1104	1140	1134	1015	1077	1116	1101	1149	1121	1148
	Max	1327	1291	1323	1323	1356	1329	1296	1341	1385	1354	1377

**Rappel des abréviations :** Hd=date d'épiaison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante, NG\_E=nombre de grains par épi, P1000G=poids de 1000 grains.

Caractère	Population	Pop. Parentale	Pop. Initiale	Le Moulon			Toulouse			Vervins		
	Génération		G0	G2	G7	G12	G2	G7	G12	G2	G7	G12
<b>HT4S</b>	<b>Moyenne</b>	<b>66.09</b>	<b>69.76</b>	<b>71.14</b>	<b>79.71</b>	<b>78.40</b>	<b>71.89</b>	<b>78.17</b>	<b>77.54</b>	<b>72.63</b>	<b>82.53</b>	<b>85.74</b>
	<i>SD</i>	<i>12.23</i>	<i>11.39</i>	<i>12.13</i>	<i>9.36</i>	<i>10.20</i>	<i>10.37</i>	<i>10.52</i>	<i>10.87</i>	<i>10.57</i>	<i>11.29</i>	<i>9.86</i>
	N	16	35	54	55	54	49	55	53	54	55	58
	Min	40	46.50	40	50.67	53.67	43.33	50.67	54.67	57.33	57	60.33
	Max	90.33	98.33	93.33	102	106.67	105.33	104	105.67	101.67	111	104.33
<b>NE4S</b>	<b>Moyenne</b>	<b>11.51</b>	<b>11.19</b>	<b>11.27</b>	<b>11.77</b>	<b>12.15</b>	<b>10.36</b>	<b>12.46</b>	<b>11.52</b>	<b>9.57</b>	<b>9.57</b>	<b>10.54</b>
	<i>SD</i>	<i>4.19</i>	<i>4.21</i>	<i>3.91</i>	<i>3.78</i>	<i>3.40</i>	<i>3.53</i>	<i>3.82</i>	<i>3.57</i>	<i>3.68</i>	<i>4.16</i>	<i>4.05</i>
	N	16	35	54	55	54	49	55	53	54	55	58
	Min	3	1.50	1	4	4.67	2	5.67	6	3.67	1	1.67
	Max	18.67	21	21	21.33	21	18.33	31	19.67	18.50	18.33	17.33
<b>NGP4S</b>	<b>Moyenne</b>	<b>257.45</b>	<b>195.96</b>	<b>178.12</b>	<b>256.45</b>	<b>218.22</b>	<b>184.82</b>	<b>327.58</b>	<b>313.77</b>	<b>187.36</b>	<b>190.37</b>	<b>228.49</b>
	<i>SD</i>	<i>231.50</i>	<i>149.42</i>	<i>129.15</i>	<i>173.14</i>	<i>180.53</i>	<i>142.09</i>	<i>175.48</i>	<i>188.50</i>	<i>163.52</i>	<i>145.86</i>	<i>182.96</i>
	N	16	35	54	55	54	49	55	53	54	55	58
	Min	1.67	8	0	10.33	4	8.70	22.77	20.33	7.50	0	0
	Max	700	634.57	569	659.33	686.33	681.67	756.90	863.67	755.33	558.33	763
<b>PGP4S</b>	<b>Moyenne</b>	<b>8.46</b>	<b>6.90</b>	<b>6.36</b>	<b>8.34</b>	<b>7.27</b>	<b>6.50</b>	<b>11.34</b>	<b>10.91</b>	<b>5.68</b>	<b>7.53</b>	<b>9.01</b>
	<i>SD</i>	<i>6.81</i>	<i>4.53</i>	<i>4.45</i>	<i>4.87</i>	<i>5.14</i>	<i>4.71</i>	<i>5.37</i>	<i>6.24</i>	<i>4.18</i>	<i>5.29</i>	<i>6.36</i>
	N	16	35	53	55	54	49	55	53	54	54	55
	Min	0.07	0.10	0.40	0.50	0.10	0.10	0.77	0.77	0.30	0.05	0.10
	Max	23.40	19.27	23.40	23.53	20.57	22.47	25.10	26.30	17.70	19.27	27.10
<b>NG_E4S</b>	<b>Moyenne</b>	<b>19.67</b>	<b>16.61</b>	<b>15.58</b>	<b>21.33</b>	<b>17.18</b>	<b>17.80</b>	<b>26.43</b>	<b>26.42</b>	<b>17.96</b>	<b>17.98</b>	<b>19.17</b>
	<i>SD</i>	<i>13.86</i>	<i>11.64</i>	<i>10.65</i>	<i>12.80</i>	<i>13.81</i>	<i>12.72</i>	<i>11.68</i>	<i>12.39</i>	<i>12.25</i>	<i>11.53</i>	<i>13.62</i>
	N	16	35	54	55	54	49	55	53	54	55	58
	Min	0.83	4.03	0	1.67	0.70	1.20	5.37	2.03	1.43	0	0
	Max	50.20	51.57	42.07	57.90	56.53	54.40	53.53	47.67	44.13	46.4	47.3
<b>P1000G4S</b>	<b>Moyenne</b>	<b>37.29</b>	<b>37.24</b>	<b>37.42</b>	<b>36.81</b>	<b>37.96</b>	<b>37.82</b>	<b>37.83</b>	<b>35.98</b>	<b>33.70</b>	<b>39.58</b>	<b>38.71</b>
	<i>SD</i>	<i>16.87</i>	<i>14.77</i>	<i>14.89</i>	<i>15.23</i>	<i>17.35</i>	<i>13.51</i>	<i>13.73</i>	<i>12.63</i>	<i>14.98</i>	<i>15.65</i>	<i>11.26</i>
	N	16	35	53	55	54	49	55	53	54	54	55
	Min	20	6.67	19.05	18.97	15.72	11.11	16.14	14.39	15.81	11.11	9.09
	Max	89.23	67.85	75.95	98.11	117.89	69.83	87.28	67.78	73.03	87.69	72.92

**Rappel des abréviations :** Hd=date d'épiaison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante, NG\_E=nombre de grains par épi, P1000G=poids de 1000 grains.

Caractère	Population	Pop. Parentale	Pop. Initiale	Le Moulon			Toulouse			Vervins		
	Génération		G0	G2	G7	G12	G2	G7	G12	G2	G7	G12
<b>HT8S</b>	<b>Moyenne</b>	<b>72.83</b>	<b>72.08</b>	<b>72.94</b>	<b>84.77</b>	<b>84.04</b>	<b>73.17</b>	<b>83.25</b>	<b>83.10</b>	<b>75.65</b>	<b>87.92</b>	<b>89.49</b>
	<i>SD</i>	<i>13.76</i>	<i>8.93</i>	<i>13.25</i>	<i>9.33</i>	<i>10.44</i>	<i>12.60</i>	<i>10.51</i>	<i>11.86</i>	<i>11.88</i>	<i>11.31</i>	<i>10.01</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	54	56	44.33	67.33	65.67	40	65.50	62	51.50	60.67	60.33
	Max	104.67	97.33	103.33	104.67	111.33	114.33	116	113.67	112	108.67	115
<b>NE8S</b>	<b>Moyenne</b>	<b>8.82</b>	<b>9.55</b>	<b>9.59</b>	<b>10.21</b>	<b>11.48</b>	<b>9.02</b>	<b>10.36</b>	<b>10.10</b>	<b>9.50</b>	<b>10.28</b>	<b>10.12</b>
	<i>SD</i>	<i>2.42</i>	<i>2.72</i>	<i>2.91</i>	<i>3.04</i>	<i>3.25</i>	<i>2.96</i>	<i>2.71</i>	<i>3.27</i>	<i>2.62</i>	<i>2.76</i>	<i>2.80</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	5.67	4.67	4	5.50	6	1	5.33	4.33	4.50	5.33	4.33
	Max	14.67	17.33	18	17.33	18.67	16.67	17	20.67	16	17.33	16.67
<b>NGP8S</b>	<b>Moyenne</b>	<b>338.00</b>	<b>399.38</b>	<b>363.78</b>	<b>439.68</b>	<b>485.84</b>	<b>379.51</b>	<b>449.18</b>	<b>450.79</b>	<b>397.31</b>	<b>398.71</b>	<b>375.14</b>
	<i>SD</i>	<i>112.49</i>	<i>116.58</i>	<i>126.16</i>	<i>172.04</i>	<i>124.89</i>	<i>148.44</i>	<i>125.03</i>	<i>154.44</i>	<i>139.41</i>	<i>120.47</i>	<i>101.76</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	174.67	202.67	80	130.67	252.33	17	195.33	225.33	137	185.67	141.33
	Max	564	681.67	625	883.67	854.07	711	762	894.33	828.33	686	650.33
<b>PGP8S</b>	<b>Moyenne</b>	<b>11.14</b>	<b>11.74</b>	<b>11.11</b>	<b>14.04</b>	<b>15.40</b>	<b>12.05</b>	<b>15.16</b>	<b>14.98</b>	<b>11.82</b>	<b>14.32</b>	<b>13.90</b>
	<i>SD</i>	<i>4.84</i>	<i>3.41</i>	<i>4.78</i>	<i>5.76</i>	<i>4.43</i>	<i>4.80</i>	<i>4.61</i>	<i>5.64</i>	<i>4.96</i>	<i>4.72</i>	<i>4.31</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	3.77	6.40	2.67	4.10	6.43	0.30	7	6.10	3.65	5.67	4.10
	Max	24.87	18.17	23.00	35.85	26.03	23.70	29.73	32.70	27.57	24.90	26.30
<b>NG_E8S</b>	<b>Moyenne</b>	<b>38.34</b>	<b>43.03</b>	<b>37.86</b>	<b>43.61</b>	<b>43.44</b>	<b>42.31</b>	<b>44.35</b>	<b>45.21</b>	<b>41.86</b>	<b>39.73</b>	<b>38.55</b>
	<i>SD</i>	<i>7.14</i>	<i>8.30</i>	<i>9.35</i>	<i>8.75</i>	<i>8.32</i>	<i>12.44</i>	<i>8.56</i>	<i>7.48</i>	<i>9.76</i>	<i>9.40</i>	<i>9.10</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	28.27	20.58	12.70	20.73	24.91	12.98	23.72	25.35	18.44	14.33	15.19
	Max	51.04	58.75	63	58.28	58.79	76.78	67.67	62.93	66.83	58.70	56.24
<b>P1000G8S</b>	<b>Moyenne</b>	<b>32.67</b>	<b>29.94</b>	<b>29.63</b>	<b>32.01</b>	<b>31.54</b>	<b>31.53</b>	<b>33.85</b>	<b>33.17</b>	<b>29.32</b>	<b>35.79</b>	<b>36.79</b>
	<i>SD</i>	<i>8.19</i>	<i>6.22</i>	<i>5.52</i>	<i>4.50</i>	<i>4.40</i>	<i>5.29</i>	<i>4.34</i>	<i>5.32</i>	<i>5.88</i>	<i>6.02</i>	<i>6.05</i>
	N	16	35	55	55	54	51	55	53	55	55	59
	Min	17.72	15.30	17.75	23.20	23.64	17.65	22.76	19.80	21.16	22.60	21.61
	Max	45.54	40.91	39.06	40.89	42.11	45.48	42.13	44.71	42.91	48.24	47.02

**Rappel des abréviations :** Hd=date d'épiaison, HT=hauteur, NE=nombre d'épis, NGP=nombre de grains par plante, PGP=poids de grains par plante, NG\_E=nombre de grains par épi, P1000G=poids de 1000 grains.

Caractère	Population	Pop.	Pop. Initiale	Le Moulon			Toulouse			Vervins		
	Génération	Parentale	G0	G2	G7	G12	G2	G7	G12	G2	G7	G12
<b>HTchp</b>	<b>Moyenne</b>	<b>87.29</b>	<b>89.24</b>	<b>91.74</b>	<b>107.67</b>	<b>104.32</b>	<b>93.33</b>	<b>101.93</b>	<b>101.33</b>	<b>98.75</b>	<b>116.98</b>	<b>118.65</b>
	<i>SD</i>	<i>17.19</i>	<i>11.91</i>	<i>15.02</i>	<i>11.48</i>	<i>9.36</i>	<i>11.51</i>	<i>12.66</i>	<i>13.40</i>	<i>14.10</i>	<i>14.85</i>	<i>11.64</i>
	N	16	35	55	55	54	51	55	53	55	54	59
	Min	60.67	67	63	79.33	77.33	74.33	77	75.33	66.67	89.67	78.67
	Max	125	116.33	129.67	130.67	131	127	140	138.67	139.33	152.33	143.67
<b>NEchp</b>	<b>Moyenne</b>	<b>4.46</b>	<b>5.11</b>	<b>5.16</b>	<b>5.42</b>	<b>5.17</b>	<b>5.20</b>	<b>4.82</b>	<b>5.44</b>	<b>4.90</b>	<b>4.93</b>	<b>4.82</b>
	<i>SD</i>	<i>1.43</i>	<i>1.26</i>	<i>1.41</i>	<i>1.43</i>	<i>1.45</i>	<i>1.36</i>	<i>1.18</i>	<i>1.39</i>	<i>1.37</i>	<i>1.18</i>	<i>1.14</i>
	N	16	35	55	55	54	51	55	53	55	54	59
	Min	2.67	2.67	2.33	3.33	2.33	2.33	3.33	2.33	2.33	2.67	2.67
	Max	7	8	9.33	10.33	9	8	8.33	9	11.67	7.67	7.33
<b>NGPchp</b>	<b>Moyenne</b>	<b>187.85</b>	<b>221.30</b>	<b>235.48</b>	<b>266.88</b>	<b>255.60</b>	<b>243.90</b>	<b>213.92</b>	<b>235.66</b>	<b>250.24</b>	<b>238.04</b>	<b>240.72</b>
	<i>SD</i>	<i>53.39</i>	<i>62.87</i>	<i>70.84</i>	<i>66.13</i>	<i>85.09</i>	<i>70.43</i>	<i>63.83</i>	<i>76.89</i>	<i>72.52</i>	<i>66.18</i>	<i>67.91</i>
	N	16	35	55	55	54	51	55	53	55	54	59
	Min	116	105	84	150.33	91.67	115.33	94	80	106.67	85.33	92
	Max	282.67	397	459	409.67	506	392.33	346.33	390	509	399	443.67
<b>PGPchp</b>	<b>Moyenne</b>	<b>8.36</b>	<b>9.52</b>	<b>9.91</b>	<b>11.50</b>	<b>10.86</b>	<b>10.62</b>	<b>9.34</b>	<b>10.40</b>	<b>10.68</b>	<b>11.25</b>	<b>11.65</b>
	<i>SD</i>	<i>2.33</i>	<i>3.07</i>	<i>3.16</i>	<i>3.10</i>	<i>3.43</i>	<i>3.17</i>	<i>2.80</i>	<i>3.36</i>	<i>3.31</i>	<i>3.60</i>	<i>3.26</i>
	N	16	35	55	55	54	51	55	53	55	54	59
	Min	5.03	4.30	3.60	6.03	3.93	4.13	4.37	2.40	4.60	3.33	3.80
	Max	13.23	17.37	18.37	19.23	21.20	17.27	15.97	17.83	21.27	20.87	20
<b>NG_Echp</b>	<b>Moyenne</b>	<b>44.14</b>	<b>43.67</b>	<b>46.24</b>	<b>49.85</b>	<b>49.15</b>	<b>46.86</b>	<b>44.66</b>	<b>43.04</b>	<b>51.52</b>	<b>48.58</b>	<b>49.65</b>
	<i>SD</i>	<i>10.66</i>	<i>8.63</i>	<i>9.29</i>	<i>7.90</i>	<i>8.41</i>	<i>6.34</i>	<i>8.44</i>	<i>8.98</i>	<i>6.59</i>	<i>7.82</i>	<i>7.69</i>
	N	16	35	55	55	54	51	55	53	55	54	59
	Min	27.40	25.50	24.33	28.93	30.10	32.23	25.17	19.39	35.90	27.63	24.77
	Max	72.23	66.83	69.43	67.47	70.35	62.30	67.80	66.77	65.87	68.17	66.17
<b>P1000Gchp</b>	<b>Moyenne</b>	<b>45.14</b>	<b>42.83</b>	<b>42.24</b>	<b>42.98</b>	<b>42.86</b>	<b>43.56</b>	<b>43.80</b>	<b>44.22</b>	<b>42.69</b>	<b>47.03</b>	<b>48.50</b>
	<i>SD</i>	<i>6.46</i>	<i>5.34</i>	<i>4.79</i>	<i>4.62</i>	<i>4.66</i>	<i>4.38</i>	<i>3.65</i>	<i>5.19</i>	<i>4.27</i>	<i>5.32</i>	<i>5.53</i>
	N	16	35	55	55	54	51	55	53	55	54	59
	Min	32.30	33.17	28.60	32.57	33.53	33.27	37.10	30.17	34.85	31.97	37.17
	Max	54.83	57.50	50.10	54.27	54.40	56.10	50.77	63	56.20	57.23	58.53