UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : SCIENCES DU VÉGÉTAL
Laboratoire d'UMR de Génétique Végétale,
Ferme du Moulon, 91190, Gif-sur-Yvette

DISCIPLINE : Génétique, epigénétique et gestion des ressources génétique

THÈSE DE DOCTORAT

soutenue le 27/06/2013

par

Abdul Rehman KHAN

Short term response of European wheat populations to contrasted agro-climatic conditions:

a genetic analysis and first step towards development of epigenetic markers in earliness gene *VRN-A1*

Directrice de thèse :	Isabelle GOLDRINGER	Directrice de recherche (INRA, UMR de Génétique Végétale, Gif sur Yvette)
Co-directeur de thèse :	Clémentine VITTE	Charge de recherche (CNRS, UMR de Génétique Végétale, Gif sur Yvette)
Co-directeur de thèse :	Jérôme ENJALBERT	Charge de recherche (INRA, UMR de Génétique Végétale, Gif sur Yvette)

Composition du jury :

Président du jury :	Michel DRON	Professeur (Université Paris Sud, Orsay)
Rapporteurs :	Valeria NEGRI	Professeur (Université degli Studi di Perugia, Perugia, Italie)
	Stéphane MAURY	Professeur (Université d'Orléans, Orléans)
Examinateurs :	Philippe BARRE	Ingénieur de recherche (INRA, Lusignan)
	Evelyne TEOULE	Maître de conférences (UPMC, Paris)
Membres invités :	Clémentine VITTE	Charge de recherche (CNRS, UMR de Génétique Végétale, Gif sur Yvette)

Abstract

Short term response of European wheat populations to contrasting agro-climatic conditions: a genetic analysis and first step towards development of epigenetic markers in earliness gene *VRN-A1*

Genetic diversity provides the raw material for evolution and adaptation of populations and species. In agrosystems, the within-population genetic diversity is of major importance: on one hand, it can provide a buffering effect against the year-to-year variation of climate or biotic pressures and on the other hand diversity serves as a resource for the population to respond to selective pressures due to specific local conditions, thus allowing for local adaptation, particularly in the case where a population is introduced into a new location. Due to its wide geographic distribution, indicating a high adaptive potential and to its socio-economic value, wheat was chosen as model crop in this study. We focused on flowering time, which is a major adaptive trait that has been involved in wheat adaptation, leading to its ability to grow over a wide range of ecological and climatic conditions.

This PhD study was designed to gain insights on the influence of within-population diversity onto the short-term response of populations to contrasting agro-climatic conditions, by studying the genetic, epigenetic and phenotypic variation. But due to the lack of prior existence of epigenetic markers, this thesis study was finally divided into two parts. In the first part, European wheat populations coming from a set of seven farmer varieties and one modern variety, grown in separate plots on seven farms distributed across Europe for three years were studied. These populations were used to study their short term response to contrasting agro-climatic conditions in Europe by analysing their phenotypic and genotypic variations. In the second part, the effect of vernalization on the DNA methylation profile of the *VRN-A1* gene was studied in winter wheat as a first step towards the development of epigenetic markers in this gene.

Results from the first part of this study revealed that conservation history of these farmer varieties strongly influenced the genetic diversity and fine genetic structure. *Ex situ* conserved farmer varieties showed low genetic diversity and simpler structure whereas *in situ* conserved farmer varieties and mixtures revealed higher level of genetic diversity and complex genetic structure. Genetic and phenotypic *spatio-temporal* differentiation depending upon the level of diversity and structural complexity of the farmer variety was observed. The traditional varieties tend to become more differentiated than the modern variety arguing in favour of the use of diverse traditional (farmer) varieties in organic and low input agriculture systems. Interestingly, a significant phenotypic differentiation for varieties with very low genetic diversity has also been observed in this study, thus indicating that other factors, such as epigenetic variation could possibly play a role in their evolution.

The second part of the study revealed that in non-vernalized conditions, *VRN-A1* is methylated in its body but not in the 5’ and 3’ ends. Comparison of vernalized and non-vernalized plants led to the identification of a region within intron one that shows significant increase in DNA methylation in response to vernalization treatment. This hypermethylation is

positively associated with *VRN-A1* expression. Although the role of this DNA methylation shift could not be investigated in the time frame of this PhD and needs further analysis, this study allowed to characterize the changes in the DNA methylation of the *VRN-A1* gene in response to cold treatment. This provides new information and sets as the first step towards the identification of possible epialleles in our populations and provides the basis for the development of markers to monitor epigenetic variability in these and other populations.

This study at large provides useful knowledge on the understanding of farmers' varieties evolutionary response to be used in the development of different breeding and conservation approaches, taking into consideration the importance of within-population diversity, to satisfactorily address the problems of organic agriculture.

Résumé

Réponse à court terme de populations de blé européen soumises à des conditions agro-climatique contrastées: analyse génétique et première étape vers le développement de marqueurs épigénétiques dans le gène de précocité de floraison *VRN-A1*

La diversité génétique est à l'origine de l'évolution et de l'adaptation des populations et des espèces. Dans les agrosystèmes, la diversité génétique intra-population est d'une importance majeure : d'une part, elle peut fournir un effet tampon contre les variations climatiques interannuelles et les stress biotiques, et d'autre part cette diversité peut permettre l'adaptation locale des populations, du fait de leur évolution sous l'effet des pressions sélectives spécifiques aux conditions locales de la région, particulièrement dans le cas d'une introduction dans un nouvel environnement. En raison de son importance socio-économique et de son aire de culture étendue, le blé a été choisi comme espèce modèle dans cette étude, en se focalisant sur l'étude de la précocité de de floraison, un caractère adaptatif majeur qui permet au blé de croître sur une large gamme de conditions écologiques et climatiques.
Ce projet de thèse a pour objet l'analyse de l'impact de la diversité intra-population sur la réponse adaptative à court terme de populations soumises à des conditions agro-climatiques contrastées, ce par l'étude des variations génétiques, épigénétiques et phénotypiques. L'absence de marqueurs épigénétiques disponible pendant la thèse a conduit à développer deux études complémentaires. Dans une première partie, sept variétés paysannes (populations conservées à la ferme) et une variété moderne ont été distribuées et cultivées pendant trois ans dans sept fermes localisées dans trois pays d'Europe, puis étudiées pour leur réponse aux différentes conditions agro-climatiques, sous l'angle de leurs variations phénotypiques et génotypiques. Dans une seconde partie, l'effet de la vernalisation sur le profil de méthylation de l'ADN du gène *VRN-A1* a été étudié, constituant une première étape vers le développement de marqueurs épigénétiques.
Les résultats de la première partie de l'étude ont révélé que l'histoire de la conservation des variétés paysannes a fortement influencé leur diversité génétique et leur structure génétique fine. Les variétés paysannes conservées *ex situ* montrent une faible diversité génétique, avec une structure génétique simple. Les variétés paysannes et les mélanges conservés *in situ* révèlent une diversité génétique plus élevée, avec une structure génétique complexe. Une différenciation *spatio-temporelle* génétique et phénotypique a été observée, en relation avec le niveau de diversité initial et avec la complexité de structure des variétés paysannes. Les variétés traditionnelles se différencient plus nettement que les variétés modernes, ce qui plaide en faveur de utilisation dans des systèmes d'agriculture biologique et à bas intrants. De façon intéressante, une différenciation phénotypique significative a été observée pour les variétés qui présentaient une diversité génétique initiale très faible, ce qui suggère que d'autres facteurs, par example épigénétiques, pourraient intervenir dans les adaptations mises en évidence.
La seconde partie de l'étude a permis de mettre en évidence un profil de méthylation intéressant de l'ADN de *VRN-A1* : sur plantes non-vernalisées, ce gène présente des niveaux élévés de méthylation dans la partie centrale du gène, mais pas en début et fin de gène. De plus, une partie du premier intron montre une augmentation significative du niveau de méthylation de l'ADN suite au traitement au froid. Ce changement de méthylation est positivement associé au niveau d'expression du gène. Si la compréhension du rôle de cette méthylation sur la régulation de *VRN-A1* nécessite des analyses complémentaires, cette étude

a permis de caractériser les modifications de méthylation de *VRN-A1* en réponse au froid et constitue une première étape vers l'identification de possibles epiallèles dans nos populations et fournit une base à la construction de marqueurs permettant de suivre la variabilité épigénétique dans différentes populations.

En conclusion, cette étude apporte des connaissances utiles pour une meilleure compréhension de l'origine et l'évolution de la diversité génétique présente dans les variétés paysannes. Ces connaissances permettront de développer des méthodes de conservation et de sélection à la ferme, en tenant compte de l'importance de la diversité intra- populations, afin de répondre aux contraintes posées par l'agriculture biologique.

SYNTHESE EN FRANÇAIS

Réponse à court terme de populations de blés européens soumises à des conditions agro-climatique contrastées: analyse génétique et développement de marqueurs épigénétiques dans le gène de précocité de floraison*VRN-A1*

La diversité génétique est à l'origine de l'évolution et de l'adaptation des populations et des espèces. Dans les agrosystèmes, la diversité génétique intra-population est d'une importance majeure :d'une part, elle peut fournir un effet tampon contre les variations climatiques interannuelles et les stress biotiques, et d'autre part cette diversité peut permettre l'adaptation locale des populations, du fait de leur évolution sous l'effet des pressions sélectives locales, spécifiques de la région, particulièrement dans le cas d'une introduction dans un nouvel environnement. Par conséquent la conservation de l'agrobiodiversité, en particulier, et de la diversité intra-population plus particulièrement, est essentielle.

Deux stratégies de conservation sont utilisées: a) la conservation *ex situ* et b) la conservation *in situ*. Dans la conservation *ex situ*, les plantes ou semences sont conservées hors de leur lieu d'origine. Il s'agit d'une methode statique dans laquelle aucune évolution face aux variations environnementales ne peut se produire. Parce qu'il y a une nécessité de remultiplier régulièrement les semences (la viabilité chute au cours de leur conservation en chambres froides), les faibles effectifs manipulés entraînent généralement une perte de diversité génétique intra-accession.

La conservation *in situ* quand à elle est définie comme la conservation plus large des écosystèmes, au travers du maintien ou du rétablissement de populations viables dans le milieu où se sont développés leurs caractères distinctifs. Pour les plantes cultivées, la conservation *in situ* correspond plus généralement à la gestion à la ferme, dans laquelle les populations hétérogènes sont cultivées dans les champs des agriculteurs, et où elles évoluent et s'adaptent aux conditions environnementales locales. Par conséquent, il s'agit d'une approche de la conservation dynamique dans laquelle les populations évoluent en permanence en fonction des conditions de l'environnement.

Traditionnellement, les agriculteurs avaient pour habitude de maintenir différentes « variétés » sur leur ferme, produisant des semences pour un certain nombre de varietés de pays (landraces), et préservant ainsi la diversité génétique à la ferme. Mais, avec la modernisation et la mécanisation de l'agriculture, au XXe siècle, en particulier dans les pays développés, les agriculteurs ont remplacé les variétés génétiquement diversifiées par des cultivars génétiquement uniformes, à haut potentiel de rendement. Les rendements de ces cultivars dépendent cependant très fortement d'une agriculture intensive, avec de gros apports d'engrais, fongicides et pesticides pour maintenir une production élevée. Mais avec l'appauvrissent des réserves de combustibles fossiles, et l'augmentation continue des prix des intrants, la durabilité de ce système conventionnel (système agricole moderne) est mise en doute. Sous l'action de différentes initiatives régionales, ou par engagement militant, de plus en plus d'agriculteurs se tournent vers une agriculture bas-intrants, ou une agriculture biologique, plus respectueuse de l'environnement. Cependant, l'absence de programmes de sélection spécialement conçus pour répondre aux besoins de l'agriculture biologique conduit à une maladaptation partielle des variétés élites à de telles réductions d'intrants. En conséquence, les agriculteurs sont de plus en plus intéressés par les anciennes variétés de pays et variétés historiques (premières variétés sélectionnées), qui ont généralement une bonne rusticité, c.à.d. une meilleure stabilité de production en situation de stress biotiques et abiotiques.

Lorsque les agriculteurs cherchent à se procurer ces variétés traditionnelles, ils s'adressent généralement aux banques de genes, ou à d'autres agriculteurs, qui pour certains ont continué à cultiver des landraces malgré la modernisation. Lors de ces échanges de semences, les agriculteurs confrontent ces variétés à de nouveaux environnements. Par conséquent, il est important de comprendre les facteurs qui influencent la réponse de ces variétés paysannes à ces nouvelles conditions environnementales. En raison de son importance socio-économique et de son aire de culture étendue, le blé a été choisi comme espèce modèle dans cette étude, en se focalisant sur l'étude de la précocité de de floraison, un caractère adaptatif majeur, qui permet au blé de croître sur une large gamme de conditions écologiques et climatiques.

Le travail de thèse avait pour objectif d'appréhender l'influence de la diversité génétique et épigénétique intra-population sur la réponse à court terme des populations soumises à des

conditions agro-climatiques contrastées. Mais en raison de l'absence de marqueurs épigénétiques disponible sur le blé au cours de la thèse, cette étude a finalement fait l'objet de deux parties.

Dans la première partie, pour étudier l'effet de la structure génétique des populations sur leur réponse à court terme à des conditions environnementales contrastées, certaines populations d'un précédent projet europeen (FSO : Farm Seed Opportunities, FP6) ont été sélectionnées, car elles fournissaient un matériel très adapté pour étudier notre question principale. Ces populations correspondent à sept landraces (populations), maintenues par sept agriculteurs européens, auxquelles s'ajoute une variété moderne de référence en agriculture biologique. Ces 8 variétés ont été cultivées dans sept exploitations agricoles (répartis à travers l'Europe), pendant trois années. Dans ce projet, nous avons caractérisé la diversité phénotypique et génétique au niveau moléculaire des populations initiales, et des populations issues des trois années de culture.

Dans un premier temps, et en l'absence d'informations préliminaires sur la structure génétique de la plupart des variétés paysannes, les données moléculaires ont été analysées au moyen d'une analyse discriminante des composantes principales et d'une description des réseaux haplotypiques. Ces analyses ont révélé que l'histoire de la conservation des variétés paysannes a fortement influencé leur diversité génétique et leur structure génétique fine. Les variétés paysannes conservées *ex situ* montrent une faible diversité génétique, avec une structure génétique simple (1 seul groupe génétique et peu d'haplotypes differents). Les variétés paysannes et les mélanges conservés *in situ* révèlent une diversité génétique plus élevée, avec une structure génétique complexe. Pour compléter les analyses, les différenciations spatio-temporelles, tant au niveau des marqueurs génétiques que des caractères phénotypiques, ont été étudiées au moyen d'AFC, arbres phylogénétiques, AMOVA, différenciation temporelle par locus et par une étude d'association entre marqueurs et données phénotypiques. L'ensemble de ces résultats montre que la différenciation spatio-temporelle génétique et phénotypique est significative, même après seulement trois années de culture, mais cette différentiation est fortement dépendante du niveau de diversité initiale et de la complexité de structurelle de la variété paysanne. Les variétés traditionnelles se différencient plus nettement que les variétés modernes, ce qui plaide en faveur de leur plus forte utilisation dans des systèmes d'agriculture biologique et à bas intrants. Fait intéressant, une différenciation phénotypique significative a été observée pour les variétés qui présentaient une diversité

génétique initiale très faible, ce qui suggère un rôle éventuel de la variation épigénétique dans les adaptations mises en évidence.

Dans la deuxième partie, l'effet de vernalisation sur le profil de méthylation de l'ADN du gène *VRN-A1* chez le blé d'hiver a été étudié, ce qui constitue une première étape vers le développement de marqueurs épigénétiques dans ce gène. Le gène *VRN-1* a été choisi parce que c'est un gène central dans la cascade de régulations de la précocité de floraison, et que c'est un déterminant clé de sensibilité à la vernalisation chez les céréales. Deux génotypes d'hiver, sensibles à la vernalisation, ont été utilisés et les profils de méthylation de l'ADN de plantes non vernalisées et vernalisées ont été comparés en utilisant un traitement au bisulfite ainsi que d'autres techniques fondées sur l'utilisation d'enzymes de restriction sensibles à la méthylation. A partir de ces expériences, nous avons démontré l'existence d'une méthylation d'ADN au sein du gène *VRN-A1*. Ce gène présente des niveaux de méthylation élévés sur sa partie centrale. En outre, une partie de l'intron 1 montre une augmentation significative de méthylation de l'ADN après vernalisation. Ce changement de méthylation est positivement associé à l'expression du gène. Cette modification de méthylation est stable au cours du développement de la plante (stabilité mitotique), mais elle est réinitialisée dans la descendance. Si la compréhension du rôle de cette methylation sur la régulation génique nécessite des analyses complémentaires, cette étude a permis d'acquérir des informations originales permettant de conduire au développement de marqueurs épigénétiques, qui peuvent être utiles par exemple pour identifier des épiallèles et suivre la variabilité épigénétique de ce gène au sein de populations.

En conclusion, cette étude apporte des connaissances utiles pour la meilleure compréhension de l'origine et l'évolution de la diversité génétique présente dans les variétés paysannes. Ces connaissances permettront de développer des méthodes de conservation et de sélection à la ferme, en tenant compte de l'importance de la diversité intra- populations, afin de répondre aux contraintes posées par l'agriculture biologique.

ACKNOWLEDGEMENTS

Thanks to Allah for His beneficence that enabled me to accomplish this task and His prophet Muhammad (SAW), who has always been the main source of inspiration and everlasting ray of hope even in the darkest of times in my life.
I would like to thank France for allowing me to stay and providing me the opportunity to learn and experience the French culture and hospitality. My heartiest gratitude to INRA especially UMR de Génétique végétale, for providing me the laboratory facilities and for allowing me utilize its resources to complete my PhD in France.
At this precise moment when I am trying to express my feelings of gratitude and appreciation for the people who were with me in this journey, words seems meaningless and insufficient.
First I want to extend my deepest and most sincere thanks to Dr. Isabelle Goldringer for supervising my PhD work and allowing me to benefit from the expertise and materials available in the lab. I want to thank you for providing me opportunities to attend various scientific conferences, meetings and workshops. I am indebted to you for your professional, social and moral support throughout my PhD.
I want to express my heartiest gratitude to Dr. Jérôme Enjalbert, my co-supervisor. I thank you for helping me and providing me the technical, professional and moral support that I really needed during my PhD. The discussions we had at different occasions, have helped me broaden my views towards scientific approaches and solving the scientific problem that come in the way. Your constructive comments and critics on experiments, data analyses, interpretation, oral presentations and write-up were always helpful. Thanks for the parties at your house which were great fun.
My deepest thanks to Dr. Clémentine Vitte, who co-supervised my PhD. I really appreciate you for your great management and organization, which I tried to follow. I know that this will help me in my future research. I thank you for your patience during my learning period as epigenetics was completely new for me and for providing me the opportunities to meet and discuss with other known researchers of the field of epigenetics.
I am thankful to Dr. Mathieu Thomas also, whose guidance and help assisted me to better orient various aspects of my work especially in the field of population genetics. Although you are working in The Netherlands but you were always available for discussions and assistance.
Thanks to Anne-Charlotte Marsollier, who helped me with the epigenetic part of my PhD.
Thanks to Nathalie Galic and Sophie Pin for your assistance in the greenhouse experiments and for all the good discussions about various cultural aspects of French people. Thanks to Pierre Rivier and Stephanie Thepot for your help with R software. On the whole, I would like to thank all the members of DEAP (Diversity, Evolution and Adaptation of Populations), our small but dynamic research team.
I would like to thank Dr. Etienne Paux, Dr. Boulos Chalhoub my thesis committee members for their guidance, help and advises which help me to work in right direction. Thanks to Dr. Jacque Shykoff for her willingness to help in any aspect regarding the administrative tasks.
Thanks to Carine Remoué, Xavier Raffoux and Celine Ridel for helping me with DNA extraction.
A special thanks to Dr. Ashwin who is a good friend and company. Thanks to all the members of UMR de Génétique Végétale especially, Delphine Sicard, Valérie Combes, Fabrice Dumas, Stéphane Nicolas, Olivier Martin, Matthieu Falque, Olivier Langella for your useful advices during the research. I would like to extend my regards to the administrative personals of our lab like Valérie Lespinas, Rozenne Le Guyader and Khalid Ouskou who helped me in the administrative matters.

I am thankful to my nation for providing funds for my stay in France through the Government of Pakistan and HEC, Pakistan. Special thanks to SFERE for facilitating our stay in France, particularly France Lamiscarre and Anna Maneta.
I would like to thank all my friends in France and in Pakistan for remembering me in their prayers. I would like to take this opportunity to thank my parents as I could not have achieved this without their support, encouragement and prayers. I want to thank my brothers Hassaan Khan and Haisam Shahzaib Khan for their prayers.

Abdul Rehman Khan

To

My parents, grandparents

and my brothers

Table of content

Chapter 1

General introduction

1.1 Biodiversity

Biodiversity can be defined as "the totality of genes, species and ecosystems in the region" (WRI, IUCN, UNEP, 1992). It can be divided into three hierarchical categories:

a) *Genetic diversity* which refers to the variation of the genome within species. It covers the variation among different populations of a species as well as variation present within a single population, and within individuals.

b) *Species diversity* which refers to the variety of species present within a region. It can be measured by assessing the number of species in an area with or without their relationships with each other (taxonomic diversity or species richness respectively).

c) Ecosystem diversity refers to the variability of ecosystems in the region but it is more difficult to measure than the other two because of problem of defining boundaries between communities – associations of species - and ecosystems. (WRI, IUCN, UNEP, 1992).

In addition to its diverse significance to the ecosystem functioning and thereby human societies (Hooper *et al.*, 2005), biodiversity serves as a way to cope with the uncertainties of highly variable environments and help ensuring the survival of life. Although, with the alarming rate of species loss, the scientists started to pay more attention towards understanding the significance of biodiversity since the 80s, the international recognition of its importance gained momentum, both politically and scientifically, since the Earth Summit in Rio in 1992 (Cardinale *et al.*, 2012). On the political front, the member nations of United Nations, acknowledging the importance of biodiversity as an essential element for sustainable development and a global asset of tremendous value to both present and future generations, agreed to take serious actions for the

conservation of biodiversity, which led to the signing of the Convention for Biological Diversity (CBD). On the scientific front, the increasing number of publications (Figure 1.1) on the different aspects related to biodiversity over the last two decades symbolizes the recognition of the importance of the matter by the scientific community and the urgent need to better understand effects of loss of biodiversity on the ecosystem functioning and in turn on human societies (Cardinale *et al.*, 2012).

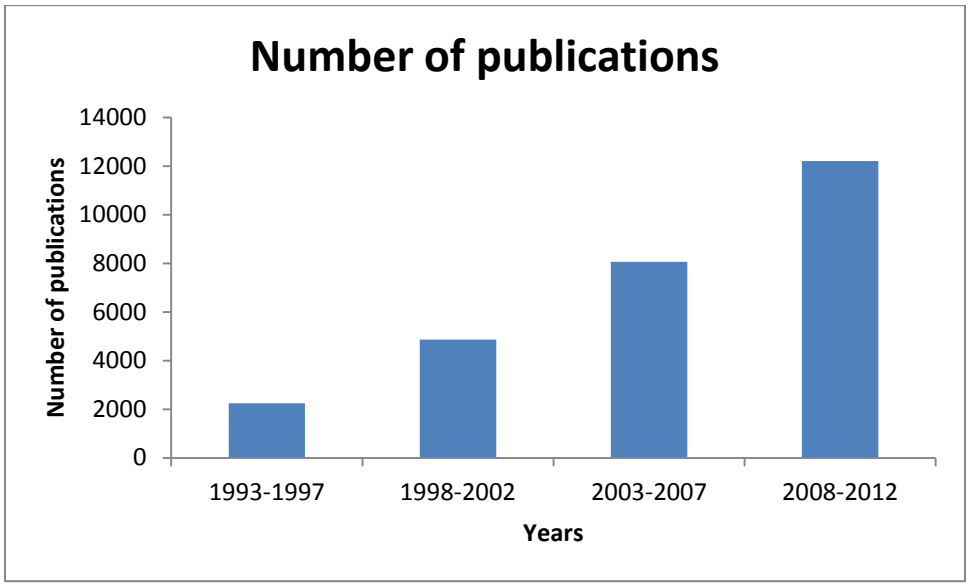


Figure 1.1: Number of publications with the word « Biodiversity » in the title in last twenty years. Searched on google Scholar

1.2 Domestication

Although the importance of human influence on the ecosystems and on biodiversity has been put into spotlight due to its drastic visible and devastating effects on environment, natural balance and biodiversity in this era of Anthropocene (Crutzen, 2002; Rockström *et al.*, 2009), the roots of human influence go deeper and long back in time. Domestication is an important step that divided the biodiversity into two groups (wild biodiversity and cultivated biodiversity) and shaped up the human societies to their present state. In the simplest terms, it is an outcome of a selection process (human selection and natural selection) which caused changes at genetic level and led to transition/transformation of wild species of plants and animals into species more adapted for cultivation and rearing. Domestication of plants started around ten thousand years

ago, when transition from hunter-gathers societies to farmers societies started (Diamond, 2002). In the case of plant domestication, hunters-gatherers selected wild plants, gathered them, brought them back to their camps and gradually started seeding and harvesting these plants. These seemingly simple activities set in motion a long term process, as initially recognized by (Darwin, 1859) and explained by (Rindos, 1984) to be the reason of many of the differences between the domesticated plants and their wild relatives, and has led to the dominance of agriculture as we know it today (Gepts *et al.*, 2012).

Domestication has social as well as ecological impacts. In the context of social impacts, domestication of plants and animals for food production purposes led to one major transformation of human societies, i.e. from hunters-gatherers society to agricultural societies. Hunters-gatherers societies were small, nomadic and had little or no division of labor. With domestication came the food surpluses, leading to larger, settled societies with division of labor, where craft specializations, arts, social hierarchies, writing, urbanization and origin of the state took place. All this, contributed towards the creation of civilizations. Domestication occurred in a few areas of the world, and at different times, therefore the societies who acquired domestication first, acquired advantages over the other societies and expanded.

Ecologically, domestication caused worldwide alteration in biodiversity and significant change in the Earth's landforms and atmosphere. Since a small number of wild populations were used for domestication, there was a strong sampling effect. This bottleneck was reported in wheat (Thuillet *et al.*, 2005) and maize (Eyre-Walker *et al.*, 1998).

1.3 **Agrobiodiversity**

After the domestication, the cultivated species went through expansion and millennia of strong evolution through mutation, migration, genetic drift, and natural and artificial selection, caused by the interaction between environment, human uses and farming practices (Purugganan and Fuller, 2009). This complex process over space and time, structured by the farmer preferences, market trends and demands, and local area adaptation due to the crop-environment interaction, has led to huge genetic differentiation which is evident at different levels of diversity between species and varying degrees of genetic structure (Haudry *et al.*, 2007). Therefore the farmers

(through traditional agriculture system), over the millennia, has given us an invaluable heritage of thousands of locally adapted genotypes of major and minor crops (Hammer and Teklu, 2008).

Agricultural biodiversity or agrobiodiversity is a component of the diversity, referring to all the diversity within and among species found in crop and domestic livestock systems, including their wild relatives, interacting species of pollinators, pests, parasites and other organisms (Wood and Lenné, 1999). The term agrobiodiversity covers all the components which are directly and indirectly involved in the production of agricultural products at genetic, species and ecological level (Jarvis *et al.,* 2007). It can also include all the mechanisms which are involved in maintaining this diversity, especially local knowledge associated with crop species that contribute to anchor a specific diversity in a specific landscape (Wood and Lenné, 1997; Jackson *et al.,* 2007).

Agrobiodiversity is very important because along with its ecological significance, it serves as an important source of raw material for breeding new varieties (Maxted, 2012) which has led to the creation of elite varieties and helped increase the food production (Huang *et al.*, 2002; Maxted, 2012). Agrobiodiversity plays an essential role in the improvement of sustainability in agricultural system and, for food security (Frison *et al.*, 2011), it provides buffering effects against stochastic or environmental changes (Gunderson, 2000; Folke *et al.*, 2004; Enjalbert *et al.*, 2011).

Traditionally, for thousands of years, farmers have cultivated crops as populations i.e. landraces, maintaining genetic diversity through cultivation of diverse landraces, but also thanks to the genetic diversity present within each landrace. In addition to their production advantages, these landraces also have social and cultural values in traditional agricultural systems, as shown by many ethnobotanists and anthropologists (Haudricourt, 1964; Elias *et al.*, 2000; Emperaire and Peroni, 2007). In this context, a landrace can be defined as dynamic populations of a cultivated species that have a historical origin, a distinct identity and lack formal crop improvement as well as often being genetically diverse, locally adapted and associated with the traditional farming systems, although to be characterized as a landrace, it does not necessarily needs to fulfill all these characteristics (Camacho Villa *et al.*, 2005; Thomas *et al.*, 2011). Therefore traditional system served as a dynamic system which allowed diversification of these crop populations,

adaptation to contrasting environmental conditions and use and maintenance of genetic diversity (Dawson and Goldringer, 2012).

In the twentieth century, agriculture went through significant changes causing a shift from traditional system to mechanized, industrialized and modernized agriculture. In this system, genetically uniform cultivars that have been commercially bred to be high yielding and "broadly adapted", replaced the genetically diverse landraces especially in developed countries of the world. Especially in the second half the twentieth century, the large increase in the use of inputs in conventional agriculture buffered and minimized the biotic and abiotic environmental variability which, otherwise, these genetically uniform varieties would have encountered (Phillips and Wolfe, 2005). So this transition led to manipulating and homogenizing the microenvironment of these homogenous modern varieties rather than using genetic diversity to buffer the environmental variability. Therefore the "broadly adapted" varieties are in fact high yielding varieties across wide geographic areas but within a narrow range of production conditions that avoid stress through heavy input usage (Ceccarelli, 1996).

The dependence on commercial varieties and on heavy inputs in the conventional agriculture system, has transformed the farming community, from producer and user of agrobiodiversity, into purely users of commercial inputs (seeds, chemical fertilizers, pesticides etc…) (Thomas *et al.*, 2011; Dawson and Goldringer, 2012). On one side, these inputs are unaffordable for the farmers of marginal areas due to their high costs, on the other hand, these diffusions of modern varieties in the traditional agricultural system has caused a dramatic decrease in the agrobiodiversity. This decline of genetic agrobiodiversity was due to two reasons. Firstly these modern varieties replaced hundreds of genetically diverse, locally adapted landraces. Secondly, many of these elite modern varieties were all very similar in their genetic constitution using similar resistance genes and genetic background (Ceccarelli, 2009; Dawson and Goldringer, 2012). In this system, farmers usually cultivate a single variety on their farms but switch frequently. Due to similarities at genetic level between modern varieties, the overall effect is reduction in diversity and tendency to monocultural landscape (Finckh, 2008) making this system vulnerable to unpredictable biotic and abiotic stresses. Two examples of vulnerability of monoculture and its consequences are southern corn leaf blight epidemic of 1970 in America which caused 15% of reduction in corn yield which was worth US$ 1 billion (Bent, 2003) and

rust attack on Cuba's sugarcane in 1979-1980 where 40% of the sugarcane area was covered by one variety and it resulted in losses worth US$500 million (FAO, 1998).

In an era of increased risk of environmental variability (Olesen *et al.*, 2011), with more variable patterns of temperature and precipitation (Olesen *et al.*, 2011) and increased costs of agricultural inputs as they are dependent on fossil fuels whose reserves are quickly depleting, the great challenge faced by the agricultural community is how to develop and improve productivity of agricultural ecosystem to alleviate poverty and ensure food security in a sustainable fashion. To meet the short term needs and achieve the long term sustainability, the role of within-variety genetic diversity (that can be found within landraces or populations) is essential (Brown and Hodgkin, 2007). It plays a major role in determining the adaptive potential of a population to new environmental conditions as well as it provides buffering capacity to the population against increasing stochastic environmental variation at both macro and micro levels. Consequently, it is essential to conserve this agricultural diversity. Two primary complementary conservation strategies, each of which includes a range of different techniques that can be implemented to achieve the aim of the strategy: *ex situ* and *in situ*. These two strategies are presented below.

1.3.1 ***Ex situ* conservation**

Ex situ conservation can be defined as the conservation of biological components outside their natural habitat (UNCED, 1992). It is a static conservation strategy which involves sampling (seeds, organs of multiplication and plants), transferring and storing of targeted taxa from the collecting site. For most of the species, seeds are stored in the genebanks. It involves the desiccation of seeds to low moisture contents and stored at low temperatures (Rao, 2004). Yet, for a number of species, predominantly important tropical and subtropical tree species, which produce recalcitrant seeds that quickly lose viability and cannot survive the desiccation, the conventional seed storage strategy cannot be applied (Engels, 2002). These seeds have to be kept moist and relatively warm. Still there are other species that are propagated vegetatively like banana, potato, sugarcane etc. Such problematic material can be conserved in the fields of genebanks and/or botanical gardens. Although it serves as a satisfactory conservation approach, it has its own disadvantages, as it is costly, susceptible to biotic and abiotic stresses and limited diversity maintenance capacity (Engelmann and Engels, 2002). In vitro conservation is another

option. With the advances in the field of biotechnology, new options like tissue culture, pollen storage, DNA banks are also available (Hammer and Teklu, 2008).

Although, the scientists had started storing more and more samples (accessions) since the first half of the twentieth century, the need to conserve the agrobiodiversity grew stronger after the "Green Revolution" because of concerns about the loss of the genetic diversity which was present in the traditional agriculture system. In response to this concern, Consultative Group on International Agricultural Research (CGIAR) started assembling the germplasm collections for the major crops. As a consequence, International Board for Plant Genetic Resources (IBPGR) was established in 1974 to coordinate the global efforts to systematically collect and conserve the world's genetic diversity. As a result of these global effects, around 7.5 million accessions of 3446 species of 612 different genera are stored ex situ around the world (FAO, 2010).

The static nature of *ex situ* conservation is a point of concern as the genetic resources conserved in the cold rooms and regenerated in the genebanks gardens are subjected to evolutionary forces which are different from those met in their original environment and these accessions could not further evolve in the environments where they have been developed after their preservation (Hammer and Teklu, 2008). In addition, the small sample size compared to original population size especially in case of heterogeneous landraces and historic varieties can also cause loss of important genetic diversity (Altieri *et al.*, 1987). During the process of regeneration (to keep the seed lot viable), there is a possibility of loss of genetic diversity and modifications in characteristics of accessions (Parzies et al., 2000; Soengas et al., 2008).

1.3.2 ***In situ conservation***

In situ conservation refers to the conservation of ecosystems, and natural habitats and the maintenance and recovery of viable populations of species in their natural surroundings and, in the case of domesticated or cultivated species, in the surroundings where they have developed their distinctive properties (UNCED, 1992). According to this definition, *in situ* conservation contains two distinct components: conservation of wild species in their natural habitat and conservation of cultivated species on-farm (in their areas of origin) (Maxted *et al.*, 2002).

The first component can be defined as "location, management and monitoring of genetic diversity in natural wild populations within defined areas designated for active, long term

conservation" (Maxted *et al.*, 1997) . The second component (on-farm conservation) is of our interest. On-farm conservation can be defined as the continuous cultivation and management of a diverse set of populations by farmers in the agroecosystems where a crop has evolved (Bellon, 1997). This maintenance of the populations in their area of origin allows these populations to evolve according to their local environmental conditions as well as local knowledge of farmers and their social norms.

1.4 Organic agriculture and its need for genetically diverse populations

In organic agriculture, the option of manipulating and homogenizing the microenvironment of crop plants through heavy inputs to limit biotic and abiotic stresses is not to be chosen is not available. Therefore, the use of heterogeneous populations enlarging phenotypic diversity to buffer the impact of environment stresses is favored. Because organic agriculture is still a small market, little interest has been shown in developing the crop varieties that could meet the needs of organic farming. Consequently, farmers are mostly using the varieties coming from conventional breeding (only the last stages of testing and seed production conducted in organic conditions;(Lammerts van Bueren *et al.*, 2010*a*) and these varieties mostly do not meet the requirements of organic agriculture (Murphy *et al.*, 2007). In addition, organic farmers are seeking for varieties with more stable robustness (*i.e.*, stable yield and quality) rather than varieties with high yield potential that require most favorable conditions (Lammerts van Bueren *et al.*, 2010*b*). This can be achieved by using heterogeneous varieties (with within-population variation) that have been reported to be more stable in terms of yield and quality in biotic and abiotic stress conditions (Wolfe, 2000; Zhu *et al.*, 2000; Finckh, 2008; Wolfe *et al.*, 2008). Landraces are known for their yield and quality stability under diverse environmental conditions (Ceccarelli, 1994), therefore organic farmers are becoming increasingly interested in their usage in organic agriculture. For this, some organic farmers have obtained seeds of these diverse landraces and historic varieties from either *ex situ* gene banks or from farmers who kept growing them even after emergence of modern agriculture. Due to the prohibition in Europe of commercially exchanging landraces that are not registered on the Official Catalogue or on the list of "conservation varieties", farmers have to grow their own seeds each year.. While *ex situ* conservation does not allow these populations to continuously evolve with changing

environmental conditions, *in situ* conservation especially on-farm conservation allows continuous evolution due to genetic diversity and farmer selection that maintains agronomic and quality characters (Louette and Smale, 2000; Berthaud *et al.*, 2001; Elias *et al.*, 2001; Smith *et al.*, 2001; Almekinders and Elings, 2001). This, in turn, helps local area adaptation of these varieties.

While such populations have a real interest in organic agriculture, only very little information about the genetic diversity found within-variety of these landraces and historic varieties grown on-farm is available as compared to *ex situ* and modern variety. The need for developing better local adaptation in organic agriculture or other contrasted agricultural conditions has been stressed in many papers, (see for example in (Murphy *et al.*, 2007), where wheat cultivars show a poor correspondence in their ranking on organic or conventional cropping system), (Gourdji *et al.*, 2013) demonstrates the absence of breeding gains in hot environment for wheat varieties selected by CIMMYT under mild climatic conditions). For these reasons, in 2007, the European Commission has funded a research program called "Farm Seed Opportunities" (FSO), which goal is to evaluate the evolution and the short term response (local adaptation) to environmental variation of historic varieties, landraces and variety mixtures currently grown by organic farmers along with a modern variety.

1.5 **Local adaptation**

Local adaptation is a process through which individuals and/or populations increase their survival chances and reproductive success in a given environment through natural selection. As the forces of natural selection vary in space, individuals interact with local environment, and ideally develop specific adaptation maximizing their fitness. Therefore, in the absence of other forces and constraints, a pattern should emerge in response to divergent selection, in which the resident genotypes have, on an average, a higher relative fitness in their local environment (habitate) than migrant genotypes (William, 1966).

The four evolutionary forces (natural selection, gene flow, mutation and genetic drift) can influence in various ways this local adaptative process, with a specific importance of gene flow quantity and quality which can hinder local selection through arrival of non-adapted individuals / genes, or foster it through input of new genetic variation (Barton and Whitlock 1997; Holt and

Gomulkiewicz, 1997). Three main mechanisms are identified to act on fitness: i) genetic adaptation and ii) phenotypic plasticity, iii) trans-generational plasticity.

Genetic adaptation can be achieved if the individuals of that population are not genetically identical (*i.e.*, if a certain level of genetic variability exists among individuals) and one part of this variability affects the traits that are linked with fitness (product of survival and reproductive success *i.e.*, number of descendants). In addition, these traits have to be under significant genetic control, in other words strongly heritable (*i.e.*, with a significant part of additive genetic variability) to be transmitted through generations. Response to local selection will thus depend on the narrow sense heritability (the ratio between the additive genetic variance and the total phenotypic variance) of these traits (Falconer and Mackay, 1996). The observation of significant phenotypic differentiation among local populations at a given trait is a clue that this trait has been involved in local adaptation and submitted to divergent selection. Depending on the architecture of the adaptive traits analyzed, specific genes or combination of genes might be selected in each environment and thus vary in frequency at a larger rate than expected under genetic drift only. Yet, adaptation may involve multiple but not necessarily substantial allele frequency changes (Le Corre and Kremer, 2012). Detecting traits involved in local adaptation and the genes underlying this selective response relies on the analysis of differentiation parameters at phenotypic, neutral markers and at the genes level but is not straightforward (Kremer and Le Corre, 2012; Le Corre and Kremer, 2012).

Phenotypic plasticity corresponds to the ability of one genotype to express different phenotypic values in different environments (DeWitt & Scheiner, 2004). This plasticity is adaptive in the sense that individuals showing a plastic response have more chance to generate offsprings (Via and Lande, 1985). Phenotypic plasticity is important, as it serves as a tool to cope with harsh environmental conditions and can reproduce thus insuring the survival of individuals, populations or even species. Because the extinction of individuals, population or species can be avoided through phenotypic plasticity, it plays an important role in population structuration, in biodiversity (genetic diversity) and in the dynamics of adaptation.

In addition to within-generation phenotypic plasticity, trans-generational effects of environment mediated by non genetic mechanisms of inheritance could influence the rate and direction of

adaptation (Jablonka, 1995; Lachmann and Jablonka, 1996; Pál and Miklós, 1999; Jablonka and Lamb, 2005; Bossdorf *et al.*, 2008; Day and Bonduriansky, 2011).

1.6 **Epigenetics and evolution**

It has long been suspected that epigenetic mechanisms, *i.e.* heritable modifications that are not caused by mutations of the DNA sequence, can explain part of the heritable phenotypic variation, and thus contribute to the evolutionary potential of natural populations (Jablonka and Lamb, 1989).

Therefore epigenetic regulations, that is, changes in gene activity that are mitotically and/or meiotically transmissible without changes in the DNA sequence (Holliday, 1994), have been subject of an increasing attention over the past years. In particular, modification of the chromatin structure allows selective reading of the genetic information contained in the genome and is now considered as a major player in development and environmental response (Roudier *et al.*, 2011). Three epigenetic mechanisms are involved in chromatin remodeling: (i) DNA methylation, (ii) incorporation of histone modification or histone variants and (iii) non-coding RNA (Rapp and Wendel, 2005). All these mechanisms, along with chromatin remodeling enzymes, play a role in modifying the chromatin states, which in turn regulate the accessibility of DNA (Kouzarides, 2007; Berger, 2007).

DNA methylation consists in the addition of a methyl group to a cytosine base leading to a 5-methylcytosine. It is evolutionarily ancient and associated with gene regulation and transposable element silencing in eukaryotes (Law and Jacobsen, 2010). Histone modification consists in post-translational covalent modifications (such as methylation, acetylation, phosphorylation or ubiquitylation) of the histone amino terminal tail and globular domain (Rapp and Wendel, 2005). These different modifications act sequentially or in combinations to obtain different transcriptional responses (Jenuwein and Allis, 2001; Berger, 2007; Lee *et al.*, 2010).

Out of these mechanisms, DNA methylation is the best understood and the most stable epigenetic mark. While it occurs mainly at CG sites in mammals, DNA methylation can occur in CG, CHG and CHH contexts (where H denotes for A, C or T) in plants (Feng *et al.*, 2010). In *Arabidopsis thaliana*, the genome-wide DNA methylation level is reported to be 24%, 6.7% and

1.7% for CG, CHG and CHH contexts, respectively. It acts differently in different regions of the genome. In transposable elements (TE), where it appears in all three contexts (CG, CHG and CHH), it is responsible for transcriptional silencing. In genes, DNA methylation is restricted to CG sites, and can influence expression (Zhang *et al.*, 2010; Law and Jacobsen, 2010). In the promoter region of the gene, presence of DNA methylation is generally negatively correlated with gene expression (Zhang *et al.*, 2006; Li *et al.*, 2012). But DNA methylation can also occur within genes (*i.e.*, away from the 5' and 3'ends of transcription units), in the so called bell-shaped CG "gene body methylation" pattern. The function of gene body methylation is still unclear (Zhang *et al.*, 2010) although it has been proposed to have evolutionary consequences (Takuno and Gaut, 2013).

In plants, cytosine methylation is regulated by two different but complementary processes which are classified as "maintenance" and "*de novo*" (Law and Jacobsen, 2010). Maintenance is the process by which pre-existing methylation patterns are maintained after DNA replication (Chen and Li, 2004). It targets symmetrical sites for which one strand remains methylated after replication. Therefore, it is restricted to CG and CHG sites. CG methylation is mainly maintained by DNA METHYLTRANSFERASE 1 (MET1), while CHG methylation is mainly maintained by CHROMOMETHYLASE 3 (CMT3). Methylation of CHHs, which are asymmetrical, needs to be reacquired *de novo* after each replication, through the action of the plant-specific RNA-dependant DNA Methylation (RdDM) pathway (Law and Jacobsen, 2010) in which small RNAs (24 nucleotides long) target the *de novo* methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) to homologous genomic loci to establish DNA methylation.

Changes in DNA methylation can influence gene expression across plant development and following stress (Bird, 2002; Zhang *et al.*, 2006, 2010; Zilberman *et al.*, 2006). This allows the expression regulation of a particular gene or set of genes, while the underlying DNA sequence remain identical (Jablonka and Raz, 2009). Most of these differences are reversible developmental effects and they are part of molecular processes underlying phenotypic plasticity in response to environmental variation (Richards *et al.*, 2010). But inheritable chromatin changes variations have also been reported (Jablonka and Raz, 2009).

Because heritability determines the potential of evolutionary changes of a trait, it is essential to determine the degree of heritability of epigenetic modifications, their impact on given ecologically important traits (Fisher, 1930*a*; Falconer and Mackay, 1996), and their role in individual adaptation to changing environment (Visser, 2008; Hoffmann and Sgrò, 2011). Several cases of naturally occurring epialleles (*i.e.*, DNA methylation alleles that are trans-generationally stable and independent of DNA sequence variation causing a visible phenotype have been described, such as the L*cyc* locus in *Linaria vulgaris* (Cubas *et al.*, 1999), a SBP-box gene in tomato (Manning *et al.*, 2006). DNA methylation natural epialleles have also been described at a larger genomic scale for species such as *A. thaliana* (Cervera *et al.*, 2002; Vaughn *et al.*, 2007), *Spartina anglica* (Salmon *et al.*, 2005) or *Populus trichocarpa* (Raj *et al.*, 2011). However, very few cases of the involvement of epigenetic variations in adaptation have been demonstrated so far, for instance in three species of *Dactylorhyzia* (***D. majalis*** s.str, ***D. traunsteineri*** s.l., and ***D. ebudensis;*** (Paun *et al.*, 2010, 2011), in *Viola cazorlensis* (Herrera and Bazaga, 2010, 2011) and in *Jatropha curcas* L. (Yi *et al.*, 2010). This is mainly because genetic and epigenetic variation covariate in most of the natural systems (Koornneef *et al.*, 2004), which hampers determining the phenotypic effect caused by the epigenetic variation *per se* (Johannes *et al.*, 2008; Richards, 2009; Richards *et al.*, 2010). Several ways have been suggested to overcome this problem. For instance, it is possible to study the consequences of chemical demethylation by using chemical agents such as 5-azacytidine (Bossdorf *et al.*, 2010), or to study epigenetic recombinant inbred lines (epiRILs) (Zhang *et al.*, 2013). However, this does not allow for natural population studies. For a limited number of species that naturally lack DNA sequence variation, such as genetically uniform clonal plant species (Gao *et al.*, 2010; Raj *et al.*, 2011), or apomict plants (Verhoeven *et al.*, 2010), genetic and epigenetic variants can be disentangled. Hence, for non clonal species, the only way to analyze the involvement of epigenetic changes in natural plant adaptation is to perform population studies of natural epialleles in parallele to classical population genetic studies. This requires to determine (i) the extent of variation in methylation patterns among individuals within a natural population; (ii) the degree to which methylation patterns affect phenotypes in this population; and (iii) the extent to which natural methylation variants are stably inherited.

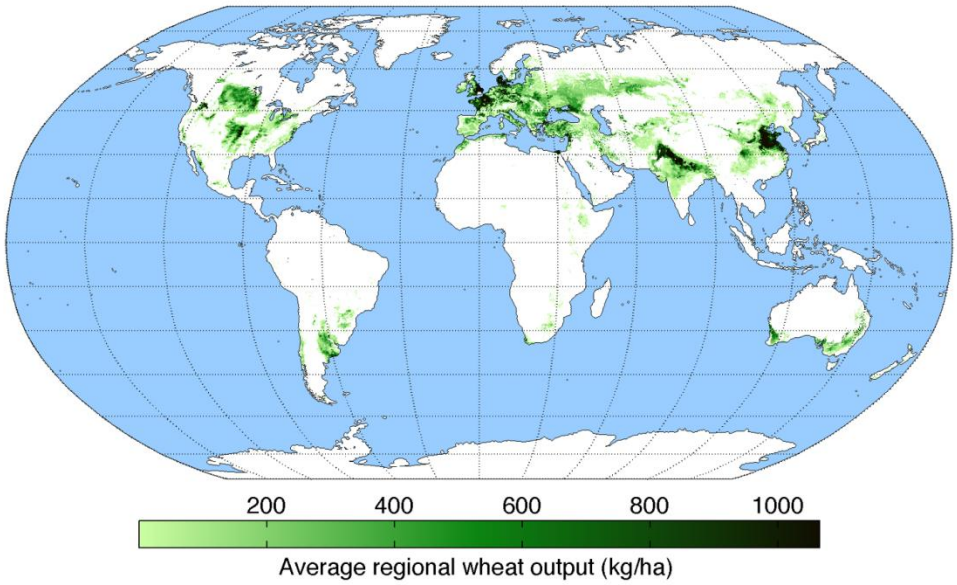


Figure 1.2: Map of wheat production across the world
(Compiled by University of Minnesota from the data of Monfreda *et al*., 2008)

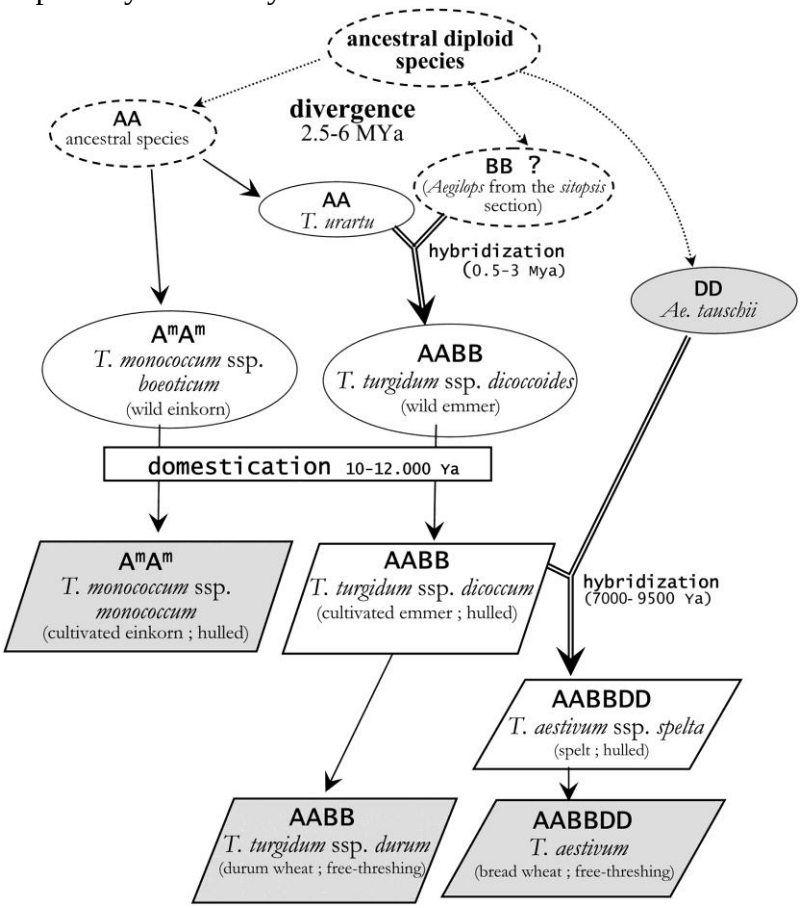


Figure 1.3: Schematic representation of the evolutionary history of wheat species (*Triticum* and *Aegilops*).Wild and domesticated species are represented in circles and squares, respectively**.**

(Chantret *et al*., 2005)

Because epigenetic variations respond to environmental changes more rapidly than the genetic variations, it could play a role in the first response involved in local adaptation. Hence, the identification of natural epialleles would be of great interest in short-term evolutionary studies of populations grown in contrasting conditions.

1.7 Wheat: An important cereal crop with high adaptive potential

Common wheat (*Triticum aestivum* L.) has a high adaptation potential to diverse environmental conditions, as shown from its geographic distribution across the globe (figure 1.2). This makes it a good biological model for studying local adaptation, especially when the populations are introduced to new environmental conditions. The importance of studying wheat is also of interest because it is one of the most important cereal crops of the world. It is ranked third after maize (*Zea mays*) and rice (*Oryza sativa*) in production and first in the area harvested (FAO, 2011). It is one of the top two cereal crops grown in the world for human consumption, along with rice although maize production is higher than both wheat and rice, the most part of the maize produced is consumed by livestock feed and or used to produce biofuel.

Wheat is one of the most ancient domesticated crops. It was domesticated in the Fertile Crescent around 10,000 years ago (Lev-Yadun *et al.*, 2000). The various species (figure 1.3) have been developed into thousands of cultivars that differ in chromosome number from the primitive diploid types, with 7 pairs of chromosomes, to hybrid allopolyploids, with 14, 21, or 28 chromosome pairs. ***T. turgidum*** (2n = 28, AABB) arose from a hybridization event that happened around 0.5 to 3 Million years ago between a diploid donor of the A genome (***T. monococcum*** ssp. ***urartu***, 2n = 14, AA) and another unknown species close to ***Ae. speltoides***, donor of the B genome (2n = 14, BB) (Figure 1.3) (Feldman et al., 1995; Blake *et al.*, 1999; Huang *et al.*, 2002). Hexaploid wheat (***T. aestivum***, 2n = 21, AABBDD) originated from an additional polyploidization event between the early domesticated tetraploid ***T. turgidum*** ssp. ***dicoccum*** and the diploid donor of the D genome, ***Ae. tauschii*** (2n = 14, DD), 7000 to 9500 years ago (Figure 1.3).

1.8 Flowering time: a major adaptive trait

The flowering time is a major adaptive trait in annual plants such as wheat and is also one of the traits involved in disease resistance. Since the beginning of agriculture, this trait has been artificially selected in crops in order to allow their cultivation in a new environment. The best documented example is perhaps that of maize which originated in Mexico tropical climate and is currently cultivated on a large part of the American continent and Europe in temperate climates. Rebourg *et al.*, (2003) showed that cultivation in temperate conditions was made possible by selecting varieties insensitive to photoperiod, flowering extremely early and are able to achieve their cycle in a short period of time, thus allowing limited culture to the warm period of the latitudes (sowing in May, harvest in October).

Flowering time is controlled by three pathways in cereals which are vernalization, photoperiod and earliness *per se* (Worland and Snape, 2001). The cultivated varieties of wheat, like most temperate cereal, can be characterized in terms of their sensitivity to cold (vernalization) and day length (photoperiod) for the initiation of flowering. Although major genes involved in environmentally influenced vernalization (*VRN-1*, *VRN-2* and *FT* genes) (Yan *et al.*, 2003, 2004, 2006) and photoperiod (*PPD-1* gene) (Beales *et al.*, 2007) have been identified, the earliness *per se*, which determines flowering independently of the environmental stimuli seems to involve a greater number of genes with weaker effects. Thereafter, QTL mapping studies have located more precisely the position of these genes on the genome of wheat (Sourdille *et al.*, 2000; Kuchel *et al.*, 2006; Hanocq *et al.*, 2007; Kamran *et al.*, 2013). Note that due to the nature of allo-hexaploid wheat genome, each gene is potentially present in three distinct homoeologous copies with little divergence. This genome architecture complicates the study of individual genes.

Due to its important role in the high adaptive potential of wheat, flowering time was selected as the adaptive trait to be investigated in this study. Thus, those genes which are associated with flowering time were included. In an era where global warming and stochastic environmental variations increase, studying such adaptive traits and how the genetic diversity of the genes controlling them, responds to this environmental variation has become more important than ever before. The study of flowering time is also interesting to understand the adaptive responses that will be implemented meet future environmental changes.

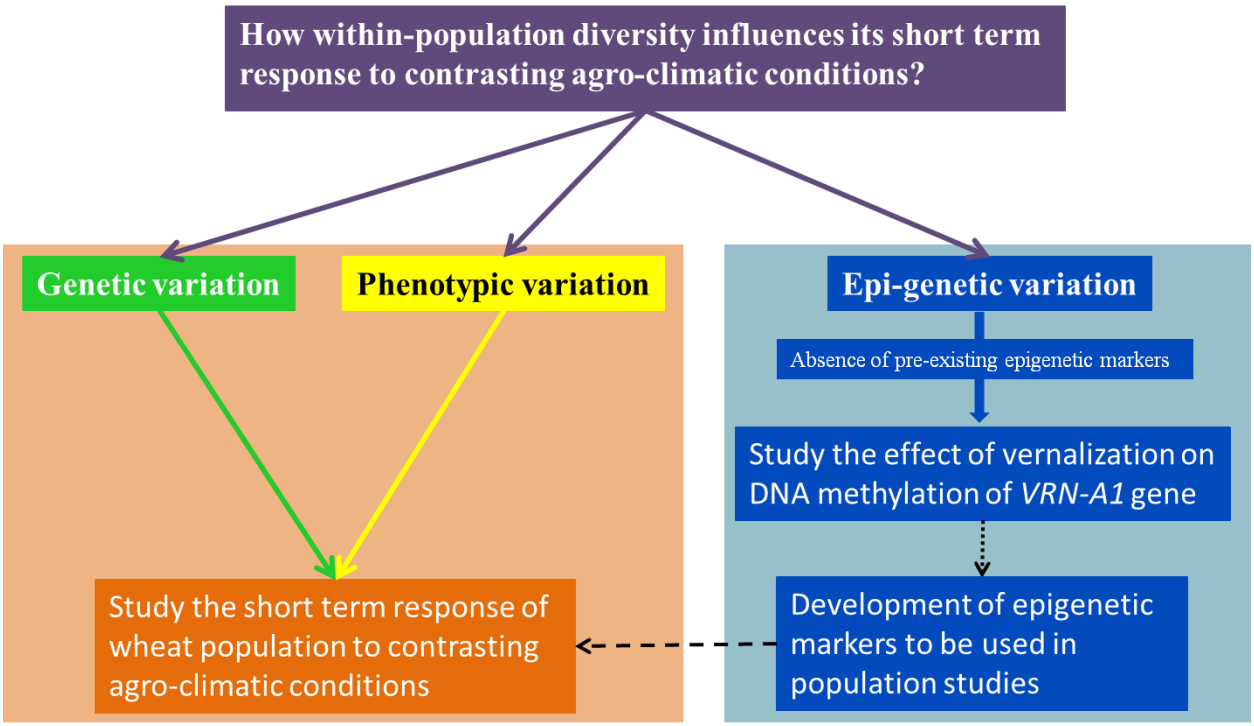


Figure 1.2: Schematic presentation of the PhD project.
Purple box: Main question; Orange box: Chapter 2; light blue box: Chapter 3 of the thesis

1.9 **Thesis plan and Objectives**

The DEAP (Diversity, Evolution and Adaptation of Populations) team is actively involved in developing various approaches for crop diversity dynamic management (on-farm conservation) and participatory plant breeding oriented towards low-input (LI) or organic agricultural systems. In this context, we are associated to a FP7 European project (SOLIBAM 2010-2014, Strategies for Organic and Low-input Integrated Breeding and Management) where Dr. I. Goldringer is leading the WP2 that focuses on "The identification of DNA and epigenetic polymorphisms for monitoring diversity evolution and for markers assistance in breeding". In this project, we focused on earliness traits (assessed through flowering time) under different environmental conditions, because these traits are key traits for plant adaptation to divergent climatic conditions and to agronomic practices such as sowing date.

As part of this SOLIBAM project, this thesis was focused on studying the role of within-population diversity in the short-term response of wheat populations grown in contrasting agro-climatic conditions. The initial aim was to study this response at the phenotyptic, genetic and epigenetic levels. But due to the absence of any prior epigenetic marker for wheat, the project was finally divided into two parts (Figure 1.2).

In the first part of this thesis (described in chapter 2), the short-term response of European wheat populations to contrasting agro-climatic conditions was studied at the phenotypic and genetic levels. This study was designed to answer two questions: (i) How do the conservation/management histories influence fine genetic structure and within-population diversity of the farmers varieties? and (ii) How do the diverse pattern of genetic structure and within-population diversity of different populations influence their short-term *spatio-temporal* differentiation response to contrasting agro-climatic conditions*?*

To address these questions, the European wheat populations from a previous EC project FSO ("Farm Seed Opportunities") were selected because these populations provided a set of diverse farmer varieties as well as a modern variety, each of which had its own conservation/management history. This set includes seven farmers' varieties (landraces, mixtures and historic varieties) collected from seven organic farmers from three countries (France, Italy and the Netherlands) and one modern variety of bread wheat (*Triticum aestivum* L.). These eight

varieties were distributed to seven farmers in autumn 2006 and each of them was grown in all locations for three years, in separate plots. For each plot, seeds from plants harvested in year n were sown in year n+1. In year 2009, i.e. after 3 generations, seed samples of each variety were collected in each farm, and a total of 48 populations were obtained.

In the time period of my PhD, I phenotyped and genotyped these populations. For the phenotypic study, these populations were grown at Le Moulon Experimental Station (UMR de Génétique Végétale, Ferme du Moulon, Gif-sur-Yvette). Genotyping was done using KASPar method (KBioscience). The genetic analysis and phenotypic analysis were performed to study the structure of these farmers' varieties, as well as the *spatio-temporal* differentiation that has occurred within this three years time span. The main objectives of this part of the thesis are:

- To study the fine genetic structure of seven farmer varieties (landraces, mixtures and historic varieties) and one modern variety to characterize the diversity maintained / developed within and among these farmers' varieties in relation to their history.
- To study the short term evolutionary effect of growing these farmers' varieties in diverse environmental and geographical conditions (7 different farms located in Italy, France and the Netherlands) for three years by evaluating the temporal and spatial differentiation at genetic and phenotypic level.

In the second part (described in Chapter 3), I studied the effect of vernalization on DNA methylation level of the *VRN-A1* gene, a central gene in the vernalization pathway. This was done to provide important basic information on the DNA methylation response of this gene, an information that is required for the development of epigenetic markers in this gene. Originally, I was planning to develop DNA methylation markers from this gene, and to apply them on the wheat population analyzed in chapter 2. However, due to the complexity of methylation analysis in hexaploid bread wheat, deciphering the DNA methylation profile of this gene took longer than originally expected. Moreover, the DNA patterns found question the use of such markers to study the epigenetic response of *VRN-A1* to cold adaptation. For these reasons, in the time frame of my PhD, I did not proceed further into the development of epigenetic markers to study

epigenetic variation at the population level as initially planned. The main objective of this part of the thesis is therefore:

- To study the DNA methylation pattern across the *VRN-A1* gene and investigate the effect of vernalization treatment on this pattern.

Chapter 2

Genetic diversity and structure of different types of wheat varieties and their short term response to contrasting environmental conditions

2.1 Introduction

Since the domestication and for thousands years, crops have been grown as populations (i.e. landraces) therefore allowing the diversification of crop varieties, adaptation to the contrasting environmental conditions, farming practices and usage, and maintenance of genetic diversity (Louette *et al.*, 1997; Elias *et al.*, 2001; Jarvis *et al.*, 2008; Dawson and Goldringer, 2012). But with the advances in plant genetics (Mendelism, F1 hybrids and pure line breeding, etc) and modernization of agriculture, these diverse historic landraces were replaced by modern homogenized varieties. This transition induced a drastic reduction of the within and among variety genetic diversity that was initially present before the industrialization of agricultural systems. This genetic erosion was reported in different studies that analysed the trends in genetic diversity over the 20th century in different crop species such as bread wheat (Roussel *et al.*, 2004, 2005), durum wheat (Maccaferri *et al.*, 2003; Thuillet *et al.*, 2005), barley (Russell *et al.*, 2000), maize (Le Clerc *et al.*, 2005), rice (Morin *et al.*, 2002) and pearl millet (vom Brocke *et al.*, 2002, 2003).

The genetic uniformity leaves a crop vulnerable to new environmental and biotic challenges and could cause serious damage to the society. One such example is corn leaf blight epidemics in America in 1970 which caused 15 % reduction in the estimated production (Bent, 2003). With the global change and thus the increased risk of environmental variability in near future (Olesen *et al.*, 2011), the within-variety genetic diversity (that can be found within landraces or population varieties) has become more important than ever before, as it plays a major role in determining the adaptive potential of a population to new environmental conditions as well as it provides a buffering capacity against increasing stochastic environmental variation.

As a consequence, conservation of genetic diversity has become crucial. The methods of conservation of agro-biodiversity or genetic resources can be classified into two types: i) *ex situ* conservation and ii) *in situ* conservation. *Ex situ* conservation involves the storage of samples (seeds, organs of multiplication and plants) in the gene banks. Mostly it includes seed storage in cold rooms but in some case (for vegetatively propagating or recalcitrant seed species) seed storage is not possible. Therefore, storage of living plants in field gene banks/ botanical gardens can be employed. However, the disadvantages of field gene banks (high maintenance costs, vulnerability to biotic and abiotic stress and limited amount of genetic variation that can be

stored) have led towards in vitro conservation methods. Recent advances in the field of biotechnology have led towards conservation of germplasm in the form of tissue culture, cryopreservation, pollen storage and DNA banks (Callow *et al.*, 1997). In 1920s and 1930s, scientists like Valvilov and Harlan became aware of the importance of crop genetic diversity, and started to collect seeds to protect the genetic diversity of traditional crops and landraces (Engels, 2002). Since then, the scientist have continued to collect and store more and more samples, called accessions, in the cold rooms of the genebanks. Currently, over 7.5 million accessions are stored *ex situ* around the world (Plucknett, 1987).

Despite this large scale of *ex situ* conservation of accessions including landraces, population varieties, historical varieties, and crop's wild relatives, all the diversity could not be collected in gene banks. In addition, due to the static nature of this conservation the accessions could not evolve after their preservation and therefore could serve very little when continuous adaptation for growing in changing environment is required (Simmonds, 1962; Henry *et al.*, 1991; Wolfe *et al.*, 2008). These constrains in the *ex situ* conservation led the scientists in 1970-1980s to come up with a complementary approach for conservation refered to *in situ* conservation where the diversity is maintained in the field (Pistorius 1997; Fowler, Hawtin, and Hodgkin 2000). Convention of Biological Biodiversity defined the *in situ* conservation as "the conservation of ecosystems and natural habitats and the maintenance and recovery of viable populations of species in their natural surroundings and in case of domesticated or cultivated species, in the surroundings where they have been developed" (UNCED, 1992). The latter case is usually considered to be on-farm conservation, which can be defined as "continuous cultivation and management of a diverse set of populations by farmer in agro-ecosystem where they have evolved" (Bellon 1997). This approach ensures the conservation of genetic diversity over time through the action of all evolutionary mechanisms (genetic drift, selection, mutation, migration). In other words, the conservation and evolution due to the natural and artificial selection goes hand in hand in this approach.

In organic farming systems, the environment is more heterogeneous both in space and time (Finckh, 2008; Wolfe *et al.*, 2008). Therefore, the cultivated species have to encounter varying environmental stresses and these stresses cannot be buffered through the use of chemical inputs. This makes the modern varieties unsuitable for these farming systems. In the absence of varieties

specially developed for organic farming, landraces or population varieties appear to be a good choice for the organic farmers. These landraces can cope with such heterogeneous environmental conditions due the presence of genetic diversity that serves as a buffering system against environmental stresses. Yet, the cultivation of once widespread landraces has been almost completely replaced by the modern varieties in the developed countries and these landraces now exist only as numerous accessions, in the cold rooms of the gene banks. The few exceptions are marginal areas or in areas where a particular landrace has a significant cultural value (Newton *et al.*, 2010). Therefore, to fulfil their needs, some organic farmers in Europe have started collecting landraces and historical varieties from the *ex situ* gene banks or from the farmers who still have these landraces and historic varieties on-farm as they had continued to cultivate these landraces even after the green revolution. They have used these landraces and historic varieties to create their own varieties (or versions of varieties) by conducting mass selection or making mixtures and letting natural selection work under their environment and agricultural practices (Newton *et al.*, 2009; Osman and Chable, 2009). Despite the recent advances in the EU legislation regarding the marketing of the so-called conservation varieties (European Commission 2008), this does not cover all types of genetically diverse population varieties and the exchange of these various types of farmers' varieties still faces legal problems. This makes it necessary that the farmers produce their own seed for each year sowing. Several farmer networks were created in the early 21st century, to facilitate the dissemination of information and exchange of knowledge and skills. These networks include Red de Semimmas in Spain, Rete Semi Rurali ub Italy and Réseau Semences Paysanne (RPS) in France (Osman and Chable, 2009; Thomas *et al.*, 2011).

Although many studies have highlighted the continuously increasing need of higher level of genetic diversity in the fields (Hajjar *et al.*, 2008), very little is known on the genetic diversity that can be found within-variety when these farmers' varieties (landraces, mixtures and historical varieties) are grown on farm compared to varieties conserved *ex situ* and to modern varieties. So, in the first part of this study, we first evaluated the genetic structure based on neutral molecular markers of seven farmer's varieties (landraces, mixtures and historical varieties) and one modern variety to characterize the genetic diversity within and among the different farmers' varieties and then compare the effect of conservation and management methods on these varieties. After deciphering the fine genetic structure of these varieties, the second part analyses the short term

Table 2.1: Description of the varieties studied (from Dawson et al 2012)

Sr. No	Name	Abbrevation	Type of variety	Farm and country of origin	Unavailable versions	Description and source of origional variety
1	Haute Loire	HL	Landrace	FFM (France)	JFB version	Landrace from the mountainous region in central France. Ex-situ conserved in French Gene bank. Under On-farm cultivation since 2004, in Northwestern France
2	Piave	PI	Landrace	GGC (Italy)	PVZ Version	Landrace from the northern Italy. Ex-situ conserved in Instituto di Genetica e Sperimentazione Agronomica di Vicenza, Italy, obtained from there regional gene bank
3	Rouge de Bordeau	RB	Historical variety	JFB (France)	PVZ Version	Historic southwestern french variety from late 1800's. Initial Sample, used in this experiment came from a farmer community in the Bourdeau region and was never conserved ex-situ
4	Redon	RD	Mixture of landraces	VVC (France)	JFB version	Farmer composed Mixture from 7 spikes, each from a different landrace accession from the Redon region from the French National Gene Bank and conserved in situ for 10 years
5	Renan	RN	Modern variety	None	JFB version	Modern french variety (INRA). The most common variety for organic farming in France
6	Solina d'Abruzzo	SO	Landrace	TDS (Italy)	None	Landrace from the mountainous region in central Italy. In continous cultivation in situ
7	Melange de Touselles	TO	Mixture of landraces	HHF (France)	JFB version	Farmer composed Mixture from 4 different ex situ conserved landraces: three T. aestivum and one T. turgidum landraces and conserved in situ for 10 years
8	Zonne hoeve	ZH	Mixture of Modern varieties	PVI (Netherlands)	JFB and PVZ versions	Mixture of two modern german varieties (Rektor and Bussard). Provided by the farmer who cultivated these two as mixture for more than 10 years in central region of the Netherlands

evolutionary effect of growing these farmer's varieties in diverse environments (seven different farms located in Italy, France and the Netherlands) for three years. In this part we evaluated the temporal and spatial differentiation at the genetic level with neutral and gene markers and at the phenotypic level to study the short term response of farmers' varieties and of a modern variety as influenced by the environmental and geographical conditions.

2.2 Material and methods

In this study, 56 populations of bread wheat were both phenotypically and genotypically analysed. For the development of these populations, seven organic farmers (four from France, two from Italy and one from the Netherlands), who already were members of seed saving associations, were contacted through the partner organizations in the European Farm Seeds Opportunities project (http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=9643492). In 2006, each farmer was asked to select a variety (which he was using in his farm at that time) on the basis of particular agronomic or quality traits, which he felt, could be of interest for other organic farmers (Serpolay *et al.*, 2011).

2.2.1 Varieties

This selected group included landraces, mixtures and historic varieties with distinct histories of conservation but grown and selected by farmers and referred to as farmer's varieties (presented in Table 2.1). Solina d'Abruzzo (SO) is a true landrace which has been continuously cultivated in its region of origin (Abruzzo, Italy) without any *ex situ* conservation. Haute Loire (HL) and Piave (PI) are two other true landraces but they were conserved in the gene bank collections and were recently obtained (last five years) by these farmers. Redon (RD) and Touselles (TO) are farmer made mixtures of several (*ex situ* conserved) landraces (several different accessions). Both of them have been under cultivation at least for the last 10 years on farm. TO is special in the sense that it comprises three *T. aestivum* components and one *T. turgidum* component. Rouge de Bordeaux (RB) is a French historic variety from late 1800's which has never been selected for genetic homogeneity. This particular population comes from a farming community near Bordeaux region. Zonnehoeve (ZH) is a mixture of two modern varieties which were cultivated,

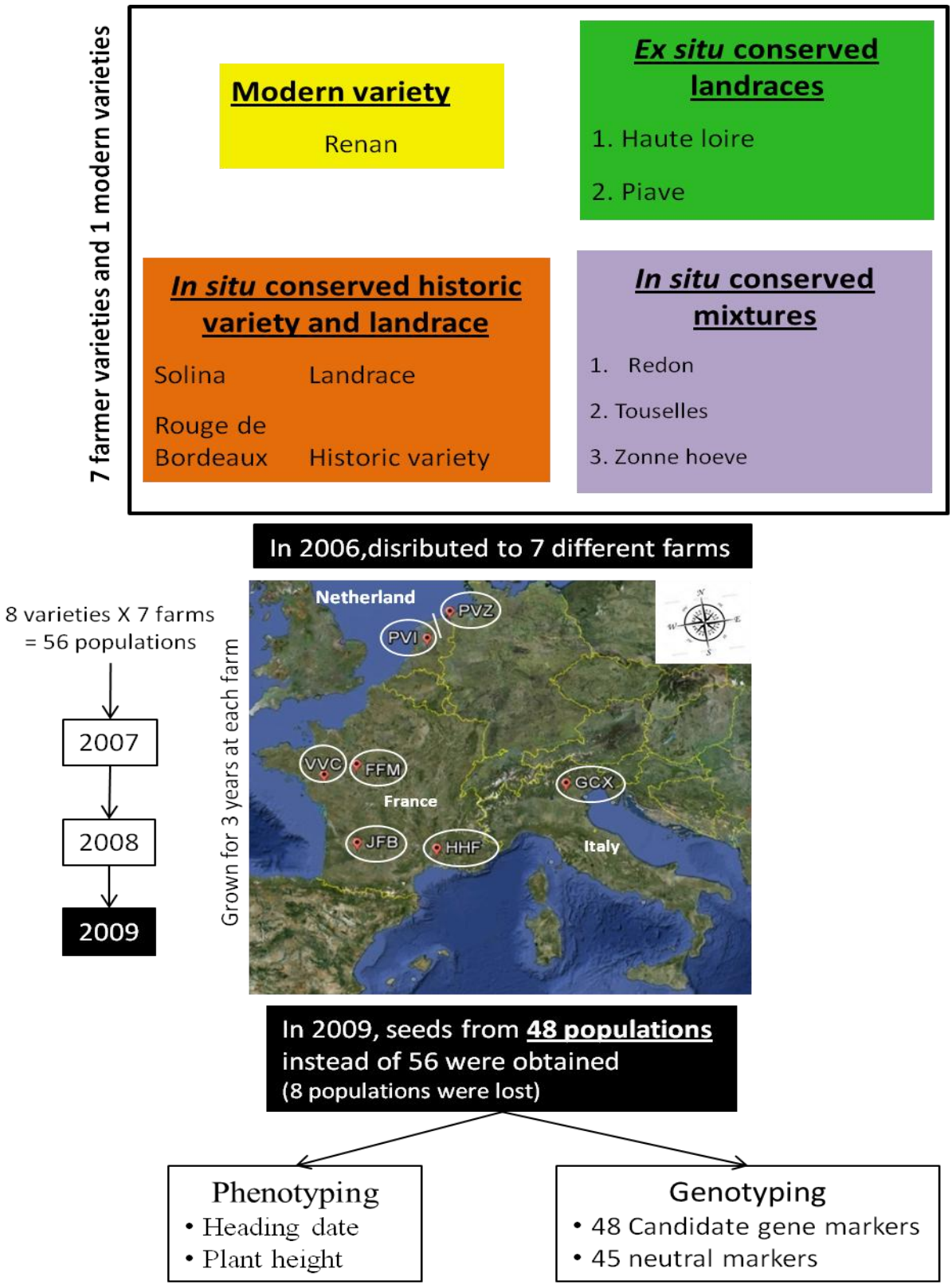


Figure 2.1: Schematic description of the experiments.
Encircled three letters are the names of the farms

harvested and replanted as a mixture for over 10 years in an organic farm in Netherland (Dawson *et al.*, 2012, 2013).

2.2.2 Experimental design

In 2006, these seven farmers varieties were sampled on the farms of origin and seeds of each variety along with a modern variety Renan (RN; currently the most widely cultivated variety in French organic agriculture) were sent to each of seven farms, mostly those of the farmers who provided varieties (figure 2.1) for sowing in 2006. For three years, the farmers multiplied each variety in a $10m^2$ plot which was separated from other plots with a distance sufficient to prevent the sample mixing at harvest (see Dawson *et al.*, 2012, 2013 for more details). The name of the farms (where these varieties were grown for three years) was used to refer the farmer's version of that variety, like variety Redon which was multiplied for three years on the farm GCX was called RDGCX.

After three years of multiplications at these farms, the seed samples were collected to be used for phenotypic and genotypic analysis. This means that after three years, there should be 56 populations (eight varieties x seven sites). Unfortunately seeds of 8 of the populations could not be recovered for different reasons (5 populations from JFB site and 3 populations from PVZ site). So the final number of samples obtained from the farmer's fields after the 3rd year (in 2009) was 48. To assess the spacio-temporal differentiation of these populations, the initial 2006 versions of each variety (8 initial varieties) were included in this experiment therefore the total number of populations became 56. These 56 populations were sown in trays on the 16th of December 2010 at the rate of 30 individuals per population. After one month in the greenhouse, these plants were transferred to the vernalization chamber for 54 days and then replanted in the open-air tunnel at Le Moulon experimental station on the 8th of February 2011 by using Randomized Complete Block Design with two replications of 15 plants/ populations of each. Since the heading date and flowering time are strongly correlated (White *et al.*, 2008) and measuring heading date in wheat (the emergence of the ear from the flag leaf), is easier than the flowering time (the emergence of the anthers from the spikelet), it is common practice in cereal experiments to record heading date as a proxy for the flowering time (Andersen *et al.*, 2004). Heading date (when half of the ear was out of the leaf sheath) and plant height at maturity were measured for each plant.

2.2.3 Molecular analysis

For each plant, total DNA was extracted from 200 mg of the 2nd leaf through DNA adsorption on Whatman Unifilter plates by following a protocol derived from the DNeasy 96 Plant kit (QIAGEN, Valencia, CA, USA).

KASPAR method (Kbioscience) was used to genotype 1650 individuals of 56 populations (~30 individuals per population) with a 96 SNP array. For this genotyping, 45 neutral markers and 48 candidate gene markers were used (Supplementary Table 2.1 and 2.2). The neutral markers have been selected from wheat 9K iSelect assay (http://malt.pw.usda.gov/t3/sandbox/wheat/termsofuse.php) (Akhunov *et al.*, 2010; Chao *et al.*, 2010). Since we expected that the varieties had undergone strong selective pressures due to their new climatic conditions as well as cultural practices, we focused mainly on climatic adaptation and looked for candidate genes associated to earliness. Thus, eight candidate genes markers (CA25, CA26, CA30, CA32,CA33, CA39, CA41, CA42) have been developed based on the previous work of INRA-Le Moulon on association mapping of flowering time genes (Bonnin *et al.*, 2008; Rhone *et al.*, 2010; Rousset *et al.*, 2011), seven (CA27, CA28, CA29, CA37, CA38, CA40 and CA44) were adapted from JIC (Beales *et al.*, 2007; Wilhelm *et al.*, 2008), one (CA31) was developed based on the published polymorphism (Su *et al.*, 2011) and the rest of them was selected from wheat 9K iSelect assay (http://malt.pw.usda.gov/t3/sandbox/wheat/termsofuse.php) (Akhunov *et al.*, 2010; Chao *et al.*, 2010).

2.2.4 Data analysis

The genetic analysis for the farmers-varieties (landraces, mixtures and historic varieties) used in this experiment was divided into two parts since there are two main objectives of this study.

2.2.4.1 *Genetic structure of farmer's varieties and landraces*

The genetic structure of the varieties was analysed based on the 41 (rest of 5 neutral markers were of bad quality so discarded) neutral markers and using the 1489 (283 individuals were rejected due to high levels of missing values) individuals from the 56 populations (initial dataset) that could be properly read.

Haplotype inference

In order to study the fine genetic structure, imputation of missing data and haplotype phase detection was carried out with PHASE (Stephens *et al.*, 2001) by using initial dataset of 41 neutral markers, which contained 1489 individuals (from the 56 populations) after discarding individuals with more than 15% of missing data (15% of individuals). Settings perform with PHASE were decided according to PHASE guideline recommendations (Garrick *et al.*, 2010) by using MR4 algorithm and a numeric experiment based on a high quality subset of 300 individuals without any missing data. The quality of the imputation was controlled by simulating different rates of missing data within the high quality data subset. Based on this work, the run was performed with 100 burns-in periods before 100 iterations, 100 number of permutations per population and a recombination rate of 0.01.

Structure analysis

To study the structure of genetic diversity, two methods, **D**iscrimenent **A**nalysis of **P**rincipal **C**omponents (DAPC) and haplotype network analysis were employed on the new data set constituted after PHASE missing data imputation for 41 neutral markers. The DAPC was done using adegenet (Jombart *et al.*, 2010), a package developed in R core (R Development Core Team, 2009). This analysis was done in two steps.

First, the script ran PCA on the the dataset and used K-means clustering of principal components to identify the k optimal groups of individuals. K-means partitions genetic variation into a between-group and a within-group component and attempts to find groups that minimize the latter. It was run with a number of group ranging from 1 to 60 and with 100000 replicates for each value of k. The Bayesian Information Criterion (BIC) derived fin each case was used to assess the best supported model, and therefore the number (k) and nature of clusters. Then, discriminant analysis was carried out on the dataset using the optimum number of groups detected. Identification of suspected migrants and outliers within populations was done on the basis of DAPC results (all the individuals present in groups other than their varietal groups were considered as suspected individuals) and then these suspected individuals were compared (with the individuals of their corresponding groups and their varietal groups) at haplotypic level to decide their status of good genotypes or migrants. Only 88 (6%) individuals were detected as

migrants. These migrants were then removed from the further analyis. DAPC analysis was done second time (on the refined dataset) to refine the grouping and check stability of the clustering.

Haplotypic analysis

Haplotype network analysis was based on the weighted adjacency matrix, which accounts for the number of differences between all pairs of haplotypes present in the dataset. Since this haplotype network analyis was done on a large number of populations (54 populations since two populations were found to be wrongly annotated and were discarded), thresholds were applied at two levels to simplify the network while retaining the maximum possible level of information. First, only the haplotypes which were detected a given number of times (*n*) or more in the dataset were selected and then, two haplotypes were connected if they were different from each other at more than *d* loci, with *d* varying from 1 to 25.

2.2.4.2 *Short term differentiation among populations within variety*

Genetic diversity

Allelic frequencies, unbiased Nei' estimate of genetic diversity (H_e) and mean observed heterozygosity (H_o) was estimated with the software GENETIX version 4.05.2 (Belkhir *et al.*, 2004). To study the genetic differentiation among different versions and generations of a variety, the data set comprising both neutral and candidate gene markers (75 markers, since 4 out of 45 neutral markers and 5 out of 46 candidate genes were of poor quality whereas 7 candidate genes were monomorphic and therefore discarded) was used for Factoral Analysis of Components (AFC) to visualize the inter-population diversity through the software GENETIX version 4.05.2 (Belkhir *et al.*, 2004). Analysis of Molecular Variance (AMOVA) was performed for 48 populations (without initial populations i.e. 2006) on neutral marker (41 markers) and candidate genes markers (34 markers) data sets seperately as well as combined (75 markers) to quantify the spatial variation (in the neutral regions and in candidate genes and overall) in terms of among varieties, among versions within variety and within populations by using Arlequin software (Excoffier *et al.*, 2005). The similarities among the populations were visualized using phylogenetic tree (for neutral markers and candidate gene markers seperately) developed with UPGMA method where the robustness of nodes was evaluated by 100 bootstraps with respect to

loci by the usage of software Population 1.2.32 and visualized by software DARwin5 (Perrier & Jacquemoud-Collet, 2006).

Population genetic effective sizes (N_e)

The population genetic effective sizes can be estimated based on the temporal variation in allele frequency among samples taken at two generations. This parameter indicates the impact of genetic drift on allele frequencies at the whole genome scale and allows to test for significant larger variation at specific locus (candidate genes) that could be due to selection.

As a first step, the allelic frequencies for the 41 neutral markers were estimated for all populations (54 populations). For each variety/landrace/mixture, the temporal variation of allelic frequencies between the initial sample of 2006 and each version of 2009 was estimated using the standardized variation of allelic frequencies at each loci ($F_{c.l}$) following Nei & Tajima (1981). Multilocus estimates F_c was estimated as the weighted average over loci of standardized variation of allelic frequency at each locus. This multilocus estimate was then used to estimate the genetic effective population size (N_e) (Waples, 1989) of each 2009 population (version) of variety as shown in the equation 1.

$$N_e = t/2(F_c - 1/(2S_x) - 1/(2S_y))$$

where t is the time between the two studied generations (t=3 in most cases), S_x and S_y are the size of samples taken from the initial population and final population respectively. In some cases, the estimated effective population size of these populations was not estimable (low sample size and limited variation in frequency between the initial and final populations led to either zero or negative values) and thus they have been replaced by the demographic population size.

Test for selection at specific loci

To detect loci exhibiting significant temporal changes as compared to the rest of the genome during the three generations, we compared the temporal variation observed at specific loci (candidate genes) to their expected distribution under genetic drift only as estimated with Ne (Goldringer & Bataillon 2004). Thus, the expected distribution of $F_{c,l}$ value for each candidate gene marker for each version (2009 populations) was derived by performing around 1015 simulations (simulations ranged from 589 to 2680) of a Wrignt-Fisher population with the

corresponding effective population size (N_e) and starting with allele frequency estimated in the corresponding initial sample.

The observed $F_{c,l}$ estimate for each marker was then compared to expected neutral distribution and an empirical *P-value* was calculated for each locus. The candidate gene markers that showed significantly higher $F_{c,l}$ than expected were considered to be potentially under selection.

Phenotypic differentiation

To study the phenotypic differentiation for heading data and plant height among different versions of each variety, analysis of variance was performed using *lm* function in R software. The model used in this analysis is given below:

$$Model\ 1 = Y_{ij} = \mu + Ver_i + \epsilon_{ij}$$

where Ver_i is the effect of different versions of a given variety. For each significant effect, adjusted means (*LS-Means*) were estimated by using the function LSD.test by using agricolie package in R core (R Development Core Team, 2009). The effect of genetic groups (groups defined by DAPC analysis) on heading date and plant height were also studied by performing analysis of variance. The model used is given below

$$Model\ 2 = Y_{ijk} = \mu + Ver_i + GG_j + \epsilon_{ijk}$$

Where GG_j is the effect of genetic groups defined by DAPC analysis.

Association of candidate genes with heading date and plant height

The association between the candidate genes polymorphisms and heading date variation (heading date data transformed into degree days from sowing) and plant height for each variety (considering all the versions of one variety at a time) except HL and RN (since both were highly conserved) was tested by considering the effect of each gene polymorphism ($Gene_i$) separately. The model used for this analysis is as follow:

$$Model\ 3 = Y_{ijk} = \mu + Gene_i + GG_j + Ver_k + \epsilon_{ijk}$$

where $Gene_i$ represents the gene effect, GG_j represent the genetic groups designated by DAPC analysis, and Ver_k represent the different versions (populations) of the same variety.

In this test, only those candidate genes that showed significant temporal differentiation for several (at least 2) of the versions within a variety were tested (CA9, CA15, CA20, CA24, CA29 and CA33 for PI variety, CA7 for RB variety, CA8 for RD variety, CA25 for SO variety, CA4 for TO variety and CA26 for ZH variety; table 2.6).

The version effect was studied more deeply by using Sample I (all the populations in a variety) as well Sample II (as using only those versions which showed the significant differentiation on that particular gene). Both type III and type I ANOVAs were applied to study the effect of genes respectively adjusted or not for the presence of the other effects.

2.3 Results

2.3.1 Preliminary results: Identification of wrong populations and of migrants in the sample

After the missing data imputation through PHASE software (see material and methods for details), Discriminant Analysis of Principal Component (DAPC) was employed on this pIMD (phase Imputed Missing Data, 1489 individuals) dataset to detect different genetic groups within the dataset. Based on the comparison of number of clusters using the Bayesian Information Criterion (BIC), the number of clusters (k) was chosen to be 25 (since the level of decrease or increase in BIC after number of clusters (k) = 25 was minimum). All the clusters/groups were assigned to one or more varieties (some groups were shared between several varieties) on the basis of the variety/varieties to which the majority of individuals of each group belonged. Most of the clusters represented only one variety (majority of individuals present in one cluster belonged to the same variety) and were assigned the corresponding names, for example almost all the individuals found in a given cluster belonged to the variety “Haute Loire” therefore this group was given the name “HL1”. Twenty-two groups corresponded mainly to one variety, while there were three clusters that shared individuals from two or more varieties and that were therefore denoted as CG1, CG2 and CG3 for Common Group (CG) 1, 2 and 3.

This analysis revealed the global structure of each variety (landrace, mixtures and modern variety) but it also identified two populations RNPVZ and TOPVZ which were not classified in their respective varietal groups. RNPVZ population belong to one specific group with individuals only found in this population. It was confirmed from *Fst* analysis that this population

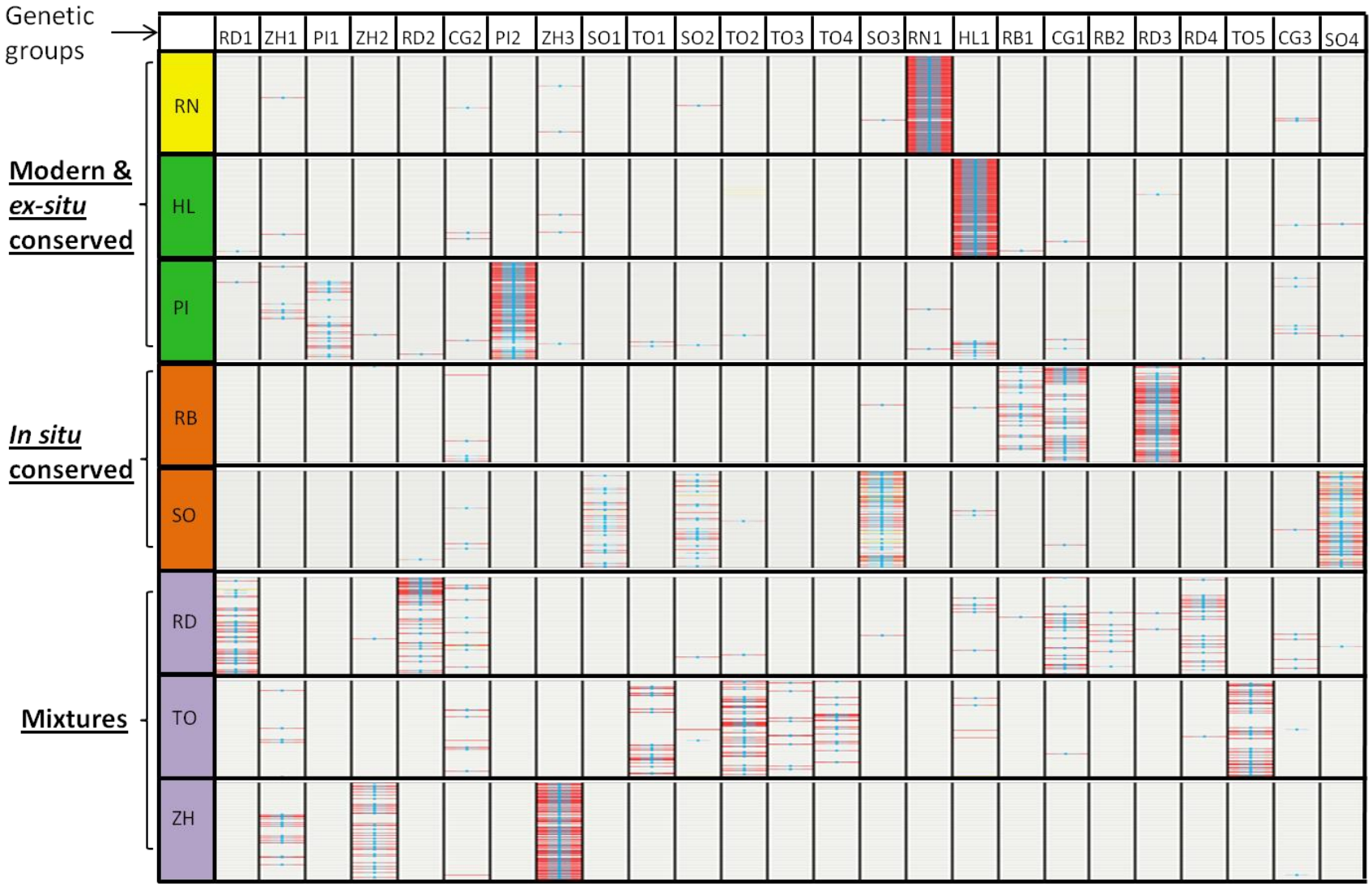


Figure 2.2 : Genetic groups defined by DAPC analysis and individuals of each variety distruibuted in these groups

was very distant from all other RN populations as well as from populations of other varieties. This was also consistent with information from the person responsible for the experiment in FSO project saying that the farmer PVZ has had difficulties in identifying some populations at harvest. Thus we concluded it was not the true RNPVZ but some other population from the farm PVZ which was mistakenly annotated as RNPVZ at harvest. TOPVZ belong to one of the groups specific for RD. The *Fst* comparison of this population with other populations of TO and RD indicated that this population was more likely RDPVZ but was mistakenly annotated as TOPVZ by the farmer. Since RNPVZ was a completely irrelevant population and TOPVZ was rather RDPVZ, both of these populations were discarded from further analysis.

DAPC analysis was also used to detect the presence of migrants (individuals issued from gene flows among the FSO experiment varieties or from external sources) in the populations. For a given variety, all the individuals that appeared at a low frequency in groups other than those assigned to this variety as well as CG2 and CG3 (as CG2 and CG3 consisted of fewer individuals and were dispersed in different varieties) were considered as potential migrants (Figure 2.2). On the basis of multilocus haplotype reconstruction analysis, these individuals were assigned as good genotypes or migrants. To minimize the risk of annotating minor genotypes as migrants, a genotype was considered as migrant, if it appeared in only one version (one population) of a variety and was closer to the other haplotypes of its assigned varietal group (DAPC assigned group) than the haplotypes belonging to the variety it was sampled from.

2.3.2 Deciphering the genetic structure of farmers varieties: middle term evolution

The genotypic data at neutral markers were analysed to study the genetic structure of the farmers' varieties (landraces, mixtures and historic varieties) and of the modern variety (Renan) used in the experiment. This is of interest because there is very little information in the literature on the genetic diversity that can be found within-variety when they are grown on farm compared to varieties conserved *ex situ* and to modern varieties.

Middle term evolutionary pattern was studied by evaluating the genetic structure (main groups) constructed by running DAPC and haplotypic networks built on the dataset of the good populations in a historical perspective. To that aim, we considered the different populations derived from a given initial variety only to characterize their genetic composition and identify

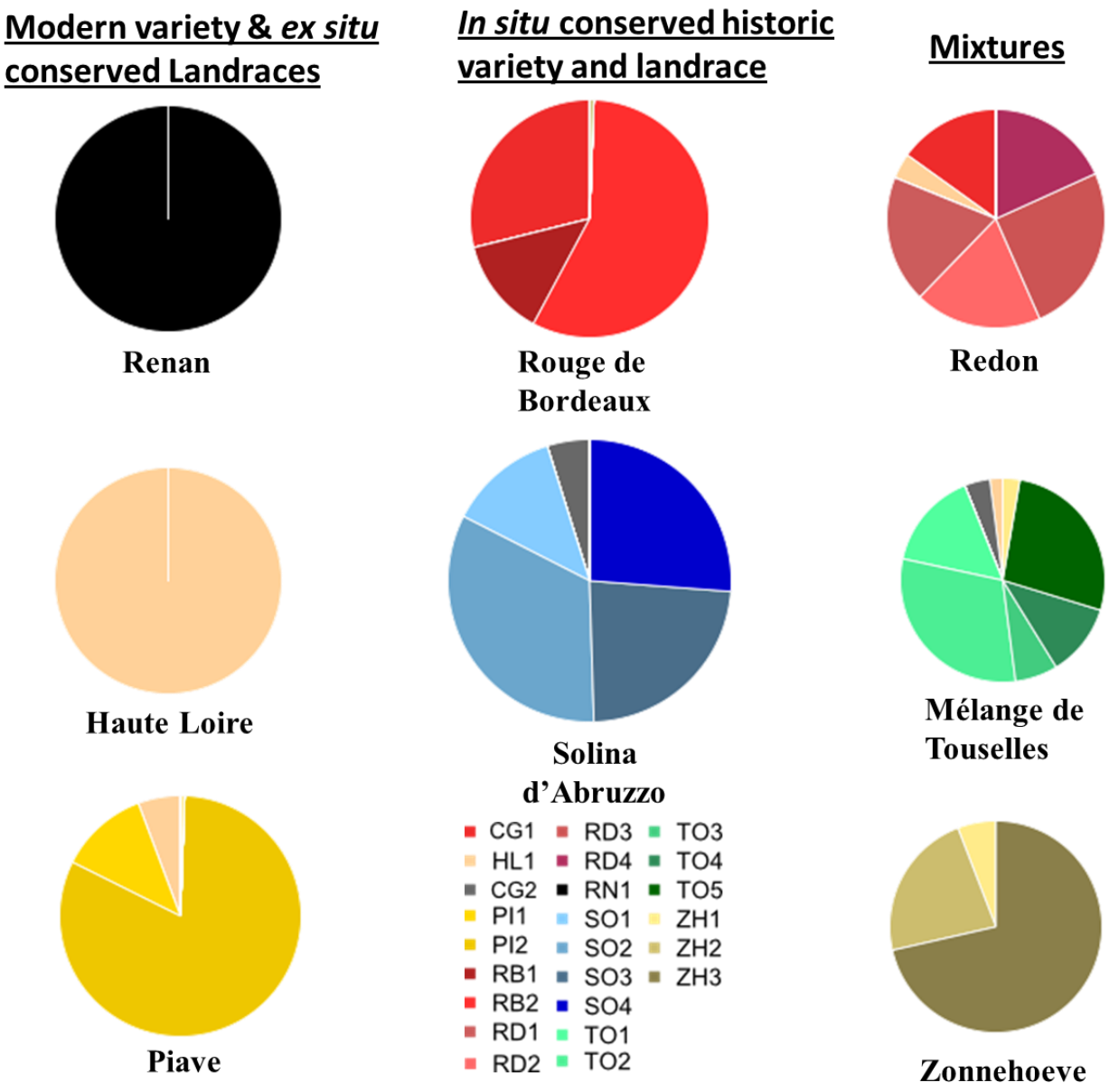


Figure 2.3: Genetic groups of each variety after discarding the migrants. Size of pie chart represents the number of individuals from each variety used in the analysis and different colors repent proportion of individuals from a particular genetic groups present in a version

potential migrant individuals, but then we studied the structure of a given variety by considering all individuals that had been assigned to the genetic group(s) of this variety, including individuals that have been detected as migrants into populations from other varieties.

The most homogeneous varieties were Renan (RN), the modern variety, and the landrace, Haute-Loire (HL). The genetic structure detected for Renan showed that all individuals were classified in a single group (Figure 2.2 and 2.3) and Nei gene diversity was very low (0.006, Table 2.2). The landrace "Haute Loire" (HL) was found to be highly homogeneous since all individuals specifically belong to a single genetic group, designed as HL1 (Figure 2.2 and 2.3) constructed by DAPC. Moreover, unbiaised Nei gene diversity of all HL individuals was found very low (0.008) indicating a low within-group genetic diversity.

Table 2.2: Diversity indexes computed for 8 populations based on 41 neutral markers

Varieties	H_e	H_o
HL	0.008	0.0006
PI	0.12	0.0027
RB	0.3351	0.002
RD	0.286	0.0043
RN	0.0058	0.0063
SO	0.1652	0.0099
TO	0.2894	0.0037
ZH	0.2198	0.0051

With H_e: unbiased Nei's estimate of genetic diversity (Nei 1978), H_o: mean observed heterozygosity

The DAPC analysis revealed that Piave consisted in two main specific groups (Figure 2.2 and 2.3). Piave was found much more diverse than HL, with unbiaised Nei diversity value 0.12 (Table 2.2). The genetic structure of the variety Rouge de Bordeaux (RB) showed three groups (Figure 2.2) *i.e.* two main specific groups (RB1 and RB2) and one shared group (CG1 was a shared group between Rouge de Bordeaux and Redon). In addition, Nei gene diversity was very high for RB (H_e = 0.335, Table 2.2). This revealed that Rouge de Bordeaux was genetically highly diverse and complex and it shared some common genetic component with "Redon"

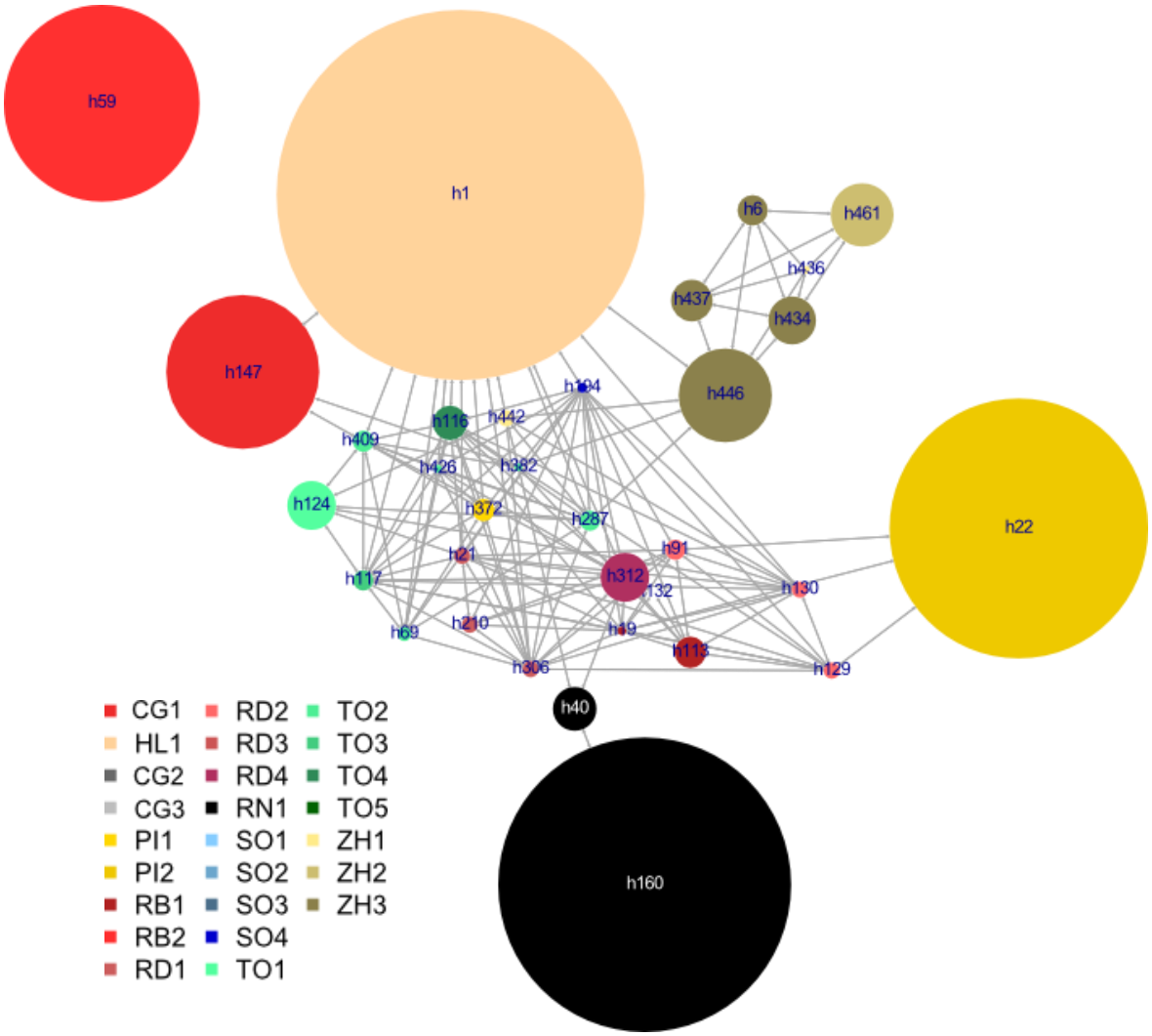


Figure 2.4: Overall haplotypic network with main haplotypes of the populations. All the haplotypes that appeared more than six times are presented and are connected if the difference between them is not more than 30. The color represents the genetic group to which the specific haplotype belongs

(landrace mixture used in this experiment). The genetic structure of "Redon" showed four main specific groups (RD1, RD2, RD3 and RD4), one shared group with Rouge de Bordeaux (CG1) and one minor group shared with Haute Loire (HL1) (Figure 2.2 and 2.3). Accordingly, Nei gene diversity was quite high (0.286; Table 2.2). The Solina d'Abruzzo (SO) genetic structure consisted in four specific groups and one shared group (CG2) with Touselles (Figure 2.3). Yet, for some individuals of this landrace, the probability, to fall in one specific group was relatively weak, but they always fall in groups specific to this landrace (or in CG2 group). In addition, the Nei gene diversity was rather high (0.165, Table 2.2). The genetic structure of the variety mixture "Touselles" (TO) showed the existence of five main specific groups (TO1, TO2, TO3, TO4 and TO5) and one group shared with SO (CG2). Accordingly, Nei within-population genetic diversity was very high (0.289; Table 2.2). Zonne Hoeve (ZH) was found to be structured into two main fixed groups, although one additional very minor group which shared some individuals of Touselle was also detected (Figure 2.2). The Nei gene diversity was found quite high (0.22; Table 2.2).

Based on the structure in 24 groups detected (since the individuals from group CG3 were confirmed as migrants) with DAPC, it is not possible to know how groups within varieties or among varieties are related to each other. To study relatedness of these genetic groups, we developed the haplotypic networks between the haplotypes (Figure 2.4). In order to simplify the haplotypic network to get a good idea of the genetic structure and relatedness of varieties, we first focused on haplotypes that were represented seven or more times in the data set (Figure 2.4, ~75% of data). In figure 2.4 two haplotypes were connected if they were different at most at 30 loci (i.e. loosely related). But in some cases, such as for Solina, very few haplotypes were present at a rate as high as seven times, therefore to get a global view, we drew the network with all haplotypes (Figure 2.5). In this network, two haplotypes were connected if they differed at less than 15 loci. Figure 2.6 is a zoom on Haute-Loire and Solina from Figure 2.5. The colors of the haplotypes correspond to those of the genetic groups they belong and that are represented in figure 2.2.

Several major haplotypes such as h160 (RN1), h59 (RB1), h22 (PI2), h147 (CG1) or h446 (ZH3) appeared little connected or peripheral while the most frequent haplotype, h1 (HL1) was highly connected to TO, SO, RD and to a lesser extend to RB and PI haplotypes. In general, haplotypes

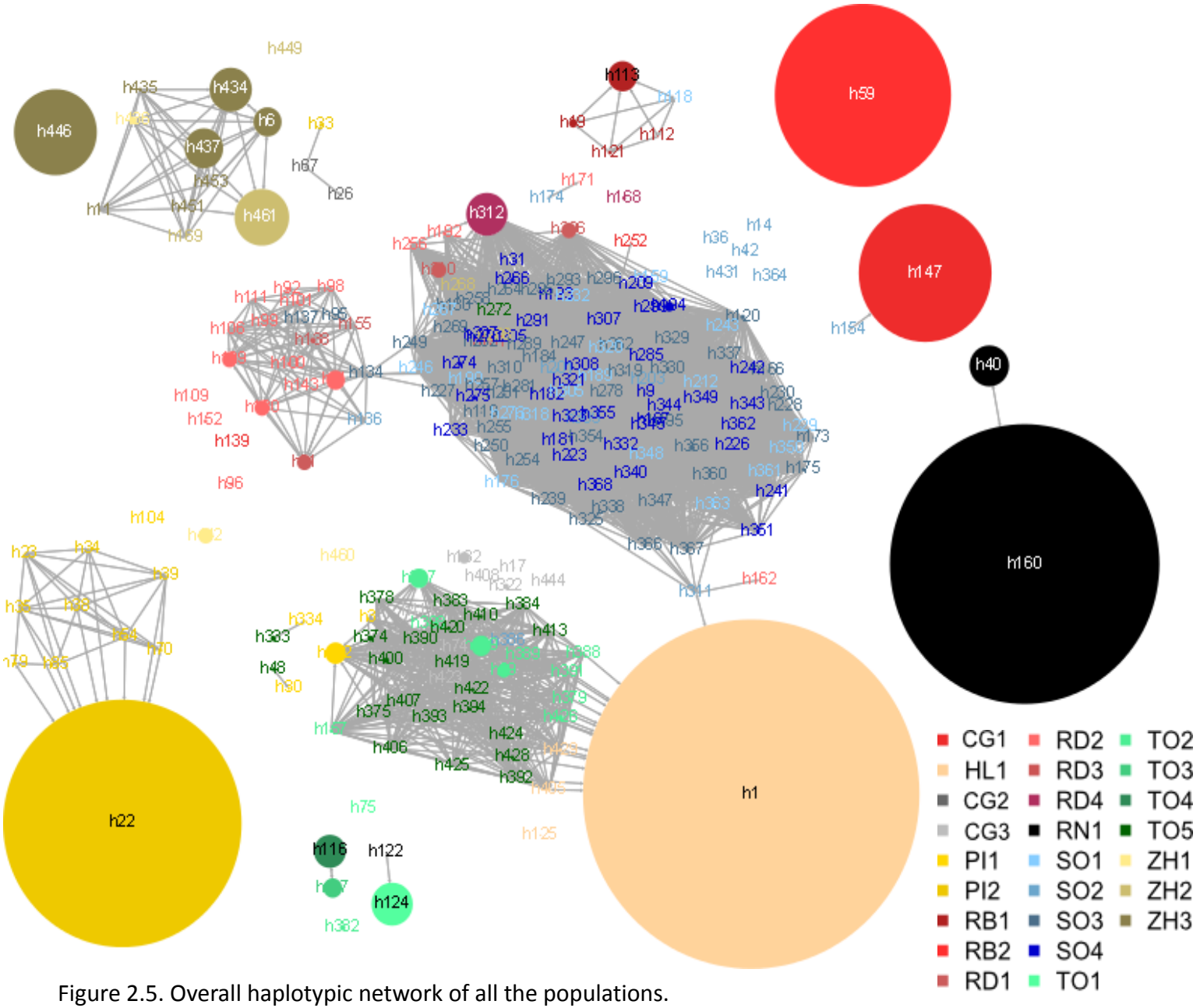


Figure 2.5. Overall haplotypic network of all the populations.
All the haplotypes that appeared more than once are presented and are connected if the difference between them is not more than 15. The color represents the genetic group to which the specific haplotype belongs

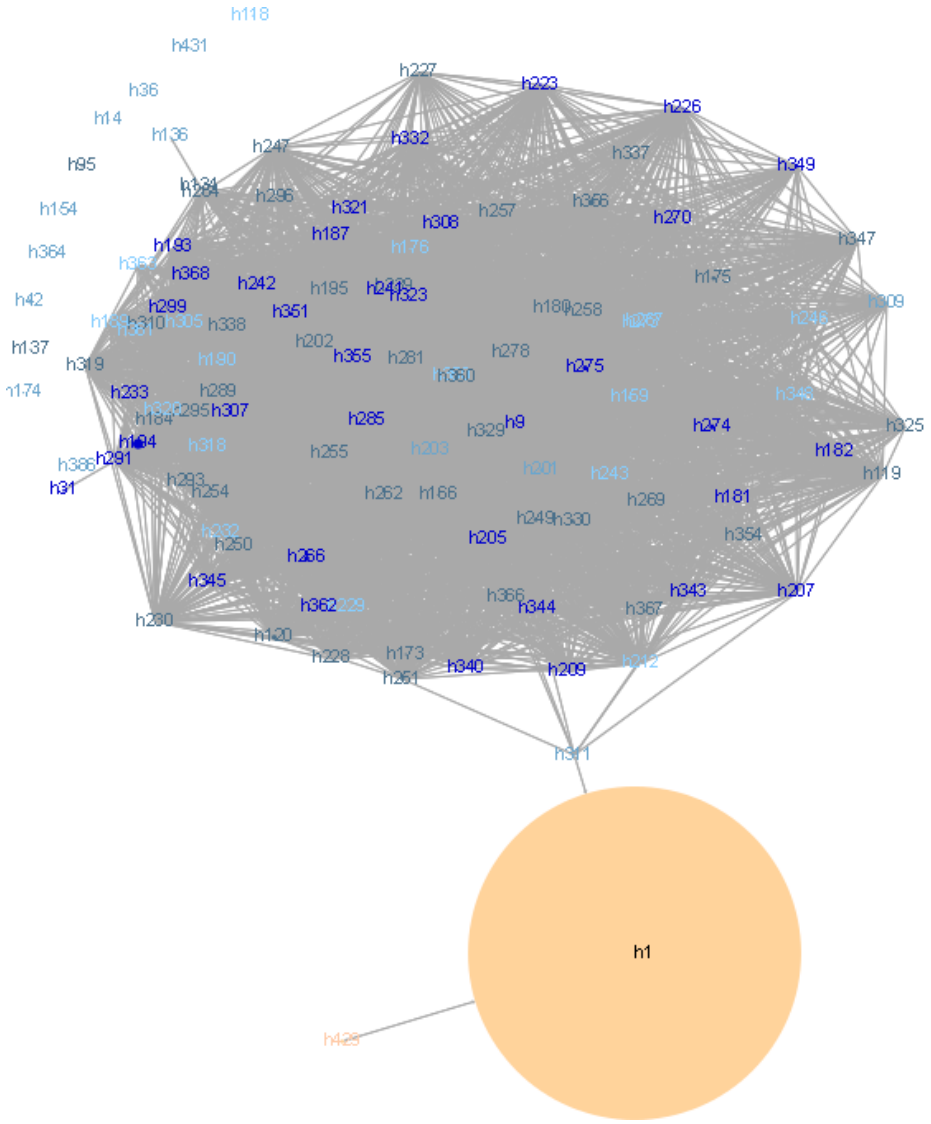


Figure 2.6: Haplotypic network of Haute loire and Solina.
All the haplotypes that appeared more than once are presented and are connected if the difference between them is not more than 15

of Haute-Loire, Solina, Touselles and Redon appeared more or less connected allogether while Renan, the Zonnehoeve components, Rouge de Bordeaux and the main Piave group were more distant from the rest of the sample. In particular, the single Renan group (RN1) was divided into two haplotypes, h160 and h40, the major haplotype, h160, being only connected to the other, which in turn was loosely connected to only two haplotypes. The main haplotypes of the two main ZH groups (h446, h437, h434, h6, h461) were quite distant from the rest of haplotypes. This suggests that the germplasm of the most recent varieties (Renan and Zonnehoeve) was genetically different from the landrace germplasm studied here. Yet, part of the more ancient germplasm (Rouge de Bordeaux, Piavé) also differed markedly from the rest of the landraces.

In general, the different haplotypes of a given genetic group and haplotypes belonging to different groups of the same variety tend to be connected at least with "loose" links (Figure 2.4 and 2.5). Yet, it was not the case for Rouge de Bordeaux, and Piavé. Haplotypes of RB1, of RB2 and of CG1 were not or only weakly connected (in particular, haplotype h59 of the RB2 group was very distant from any others) and haplotype h372 of PI1 group was weakly connected to the main Piavé haplotype, h22 (PI2 group) (Figure 2.5).

In the figure 2.6, only one haplotype representing "Solina" was present (h194) because the variety is composed of a multitude of haplotypes that were found at low frequency. This specific behaviour or genetic structure made Solina an interesting case study. When all the haplotypes appearing more than once were used, and the haplotypic linkage threshold of 15 was applied, almost all haplotypes of the four genetic groups were highly connected with each others showing no clear structure (Figure 2.6). In comparison to the highly fixed Haute Loire (one haplotype representing almost all the individuals), Solina revealed a very large number of haplotypes (148 haplotypes) and most of these haplotypes were well connected to each other (figure 2.6), revealing that these haplotypes were highly related to each other with differences at only a few number of loci.

These results show that these farmers' varieties (landraces, mixtures and historic varieties) present diverse and contrasted genetic structures ranging from the highly conserved landrace Haute Loire and modern variety Renan, to the diverse historical variety Rouge de Bordeaux composed of few distinct groups, mixtures Redon and Touselles composed of more heterogeneous groups, to the most diversified landrace Solina but with highly connected

haplotypes. This indicated that in addition to natural selection, the manner in which humans historically handled these varieties strongly influenced their structures.

2.3.3 Migrants and varietal relatedness

As described earlier, migrants (individuals that migrated from their respective variety to another variety used in this experiment) were identified through the DAPC and multilocus haplotype reconstruction analysis based on the 41 neutral markers (Table 2.3). The overall percentage of migrants in the dataset was found to be 7.3%. Migrant comparison at version level revealed that the versions from PVI contained the highest percentage of migrants (17.12%) followed by PVZ version (12.66%), For other farmers the average rate of migrant per version was less than 10%. At the varietal level, Piave showed the higest level of migrants (15.8%) followed by Redon (10.6%) which indicated that there was not link with the type of variety and its heterogeneity. While it might be more difficult for farmers to identify migrants in a highly heterogeneous population such as Redon, Piavé was phenotypically homogeneous. This level of migrants seems rather high but it should be kept in mind that these populations were cultivated in small plots which is not a normal practice for farmers.

Some individuals which belonged to different varietal group than their original varietal group, but appeared in more than one population of that variety therefore were not identified as migrants since we assumed migration occured on farm during harvest or sowing or by cross-pollination and thus is expected to depend on the farm rather than on the variety. One such case is HL group HL1. Many individuals from more than one version of Piave, Redon and Touselles were in this group. The haplotypic reconstruction analysis revealed that all the individuals from PIPVI (5 individuals), PIVVC (4 individuals), RDGCX (5 individuals), RDHHF (1individual) and TOFFM (2 individuals) in HL1 group belong to same haplotype (main haplotype of Haute Loire) while individuals from TOHHF (2 individuals) in HL1 group are different at 3 loci from main haplotype of Haute Loire. The presence of this haplotype in diverse populations like Redon and Touselles and its complete absence from the modern variety Renan (known to be highly fixed) advocates for the hypothesis that this haplotype was the part of initial diversity of these varieties.

2.3.4 *Spatio-temporal* evolution of farmers' varieties and landraces

2.3.4.1 Overall differentiation among varieties and versions at neutral and candidate genes markers

AFC analysis of these 54 populations was done separately for 41 neutral markers and 34 candidate gene markers, to have a general idea about the genetic proximity between the populations and to compare the pattern obtained with candidate genes and with neutral markers. The projection of first three components (figure 2.7A and B), which explained 71.3% and 71% of the total variance for neutral and candidate gene markers respectively, provided a good picture of the genetic diversity among the populations in both time and space, in terms of varieties and versions within each variety. Comparison between the two AFC showed that the different versions of each variety were more distant from each other for neutral markers than for candidate gene markers.

The different versions of Haute Loire appeared to be highly similar both for neutral markers and for candidate gene markers. They appeared close to Touselles, Rouge de Bordeaux and Redon at the candidate genes level but seemed distinct from these varieties at neutral markers, although this distinctness was mainly due to the 3rd axis which represented only 15% of the total variance. This indicates that the Haute Loire shared similar candidate gene make up with these three varieties in a relatively different neutral background. Piave showed a high level of divergence from the rest of varieties in particular for the candidate gene markers, but while the different versions were quite similar for the candidate genes, they differed to some extend from each other at neutral markers. Unlike in the haplotype network, Rouge de Bordeaux clustered with Redon and Touselles, forming a cluster named as RRT cluster (Rouge de Bordeaux, Redon and Touselles cluster). These varieties showed differentiation among the different versions. Interestingly the versions of each of these varieties (as well as the Haute-Loire versions) appeared to be closer at the candidate gene markers than at neutral markers giving an impression of having similar genetic makeup at candidate genes in a relatively heterogeneous neutral background. Solina showed very low differentiation among its versions both at neutral and candidate gene markers. It appeared to be slightly genetically related to RRT cluster for both type of markers. Piave was quite distant from the others at the candidate genes markers with quite a lot of variation among versions at both levels. Renan appeared to be genetically distinct

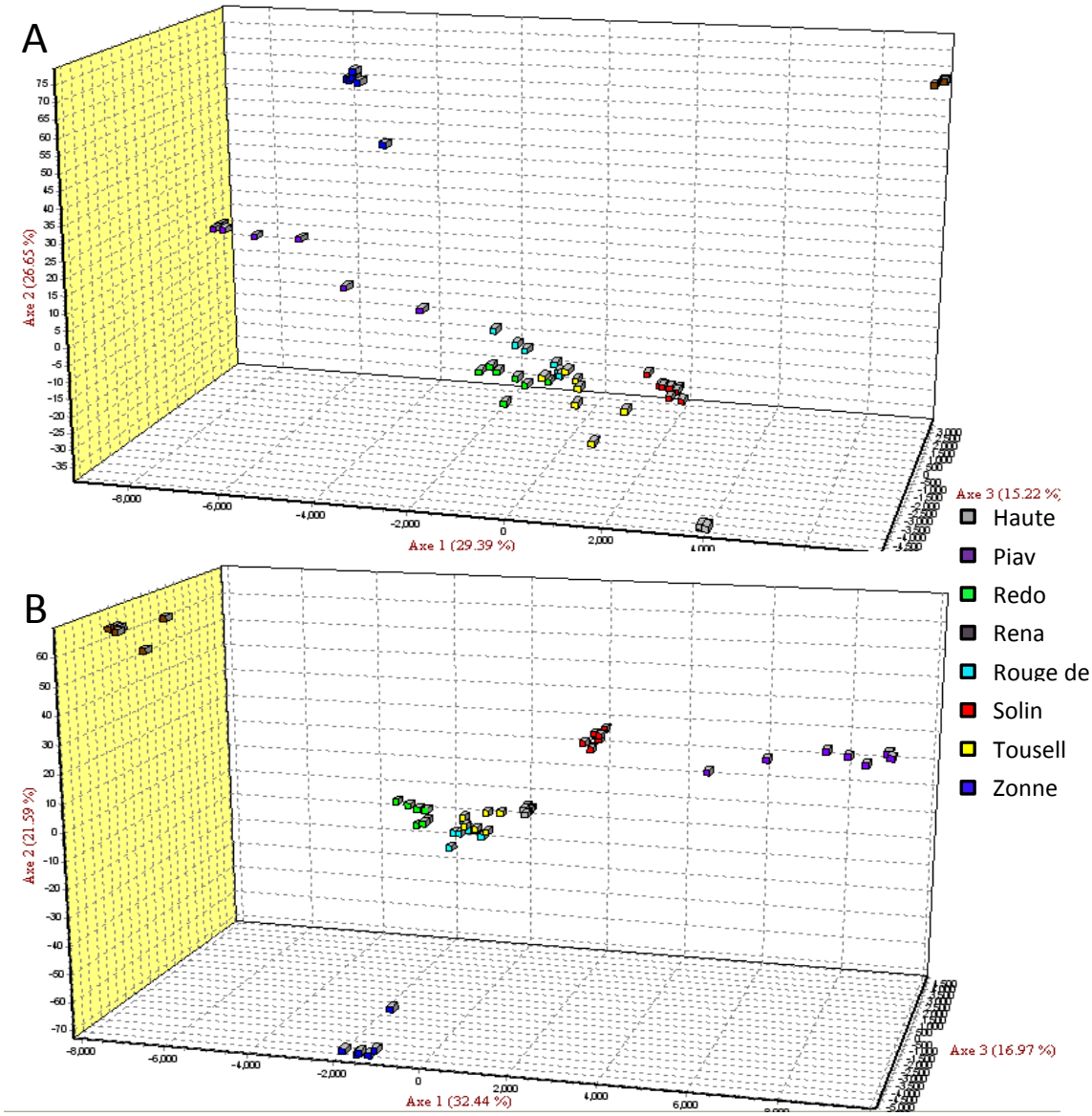


Figure 2.7: Graphical projection of 54 populations as defined by the first three axis of AFC. A) for neutral markers and B) for candidate gene markers. Colors in the boxes represent the different varieties.

from the rest and highly conserved among its different versions for neutral markers but showed some differentiation between its versions at candidate gene markers. Zonnehove also appeared to be distinct from the rest of the varieties and show differentiation at GC version more strongly than other versions at both neutral and candidate gene markers. From this AFC analysis, it can be assumed that although the variance mainly occurs among varieties, there is a certain level of variance among different versions within variety.

2.3.4.2 Spatial differentiation among versions at neutral and candidate genes markers

To quantify the level of spatial differentiation at different levels (among varieties, among versions within variety and within population) and to test the significance of this variation, analysis of molecular variance (AMOVA) was performed on the data set that excluded the initial populations (2006 populations) on neutral markers (41 markers) and on candidate gene markers (34 markers) separately as well as combined (75 markers). In the AMOVA (table 2.4) the variation among varieties was the largest with 62.36 % of the total observed variation, then the variation within populations reached 34.54% and the lowest was the among version within variety variation (3.09%).

The structure of variation was quite similar for candidate genes on one side and for neutral markers on the other (Table 2.4). While the variation among varieties was slightly higher in candidate gene markers than in neutral markers (respectively 65.12% and 61.31%), within population variation was lower in candidate gene markers than in neutral markers (respectively 31.79% and 35.61%) but interestingly the level of variation among different versions within a variety remained constant in candidate genes and neutral markers. Thus, although more than 95% of the variation observed was due to varietal differences and the within population diversity, there was a small but significant portion (3.09%) of variation due to the changes under the different agro-climatic conditions encountered by the varieties during their cultivation at geographically different sites or regions.

The phylogenetic tree for neutral markers and candidate gene markers was constructed to understand the relatedness between different versions of same variety and of different varieties (Figure 2.8). While most often the different versions of each variety clustered with a rather good reliability (high bootstrape values), the clustering among varieties was quite poor. In addition,

Table 2.3: Percentage of migrants detected within each population per farm per variety

Variety \ Version	FM	GC	HF	JFB	PVI	PVZ	VC	%age per Variety
HL	0.00	10.00	6.90		14.29	16.00	10.71	9.60
PI	6.67	3.45	6.67	0.00	41.38		14.81	15.80
RB	0.00	0.00	14.81		17.48		0.00	2.00
RD	0.00	11.54	6.67		21.74	10.71	6.67	10.60
RN	3.33	6.67	6.90		3.85		0.00	5.60
SO	0.00	0.00	4.17		14.81	11.54	3.33	4.70
TO	0.00	0.00		10.71	4.55		7.14	3.90
ZH	0.00	0.00	17.86	0.00			6.90	5.00
%age per Version	1.79	4.03	7.97	3.57	17.12	12.66	6.19	

Table 2.4: AMOVA for 46 populations (2009 versions of each variety) based on all 75 markers (neutral and candidate gene markers combined), 34 candidate gene markers and 41 neutral markers

Source of variation	d.f.	Overall (75 markers)				Candidate genes (34 markers)				Neutral markers (41 markers)			
		Sum of squares	Variance components		Percentage of variation	Sum of squares	Variance components		Percentage of variation	Sum of squares	Variance components		Percentage of variation
Among Varieties	7	16605.93	8.39015	Va	62.36	5016.778	2.53631	Va	65.12	11599.95	5.85956	Va	61.31
Among versions within varieties	38	939.546	0.41626	Vb	3.09	267.641	0.12035	Vb	3.09	670.077	0.29501	Vb	3.09
Within populations	2196	10204.76	4.64698	Vc	34.54	2719.498	1.23839	Vc	31.79	7473.634	3.40329	Vc	35.61
Total	2241	27750.24	13.45339			8003.917	3.89504			19743.66	9.55786		

although the fine clustering appeared rather similar at neutral markers and at candidate genes, the grouping of varieties was quite different except that in both cases, Renan was very distant from the other varieties, and Rouge de Bordeaux and Redon were rather close.

Looking more finely in the grouping of the versions, we see that except Rouge de Bordeaux at neutral markers, which split into two groups, one with Redon versions, one specific, all the version of each variety clustered together with eitheir type of markers, indicating that whatever the level of within-variety initial diversity and spatial differentiation among versions, the varieties stayed distinct from each other.

All versions of Haute Loire were very closely related for both neutral and candidate genes marker which is consistent with their high genetical homogeneity (a major haplotype in all versions). Renan showed strong relatedness among different versions including the initial RN population at both types of markers. Interestingly HHF version showed some distinctness from the rest only in the candidate genes indicating some specific differentiation. The versions of Solina also showed high level of relatedness in both candidate genes and neutral regions, although the JFB version in neutral regions appeared to be slightly distinct from rest of the SO versions including the initial SO population. Piave appeared more diversified. GCX, HHF and JFB versions showed low differentiation from initial population while VVC, FFM and PVI versions showed stronger differentiation from initial population at both neutral and candidate gene markers. In Rouge de Bordeaux at neutral markers, some versions clustered with the initial population and were closer to Redon, while the others were more distant. This might be linked to the frequency of the CG1 group in the different samples. Redon also presented varying level of relatedness among different versions (including initial population) and showed relatively higher level of distinctness in neutral markers than in candidate gene markers. Some versions of Touselles presented certain level of distinctness from each other but which was not consistent in candidate genes and neutral markers. Zonnehoeve versions showed the same pattern relatedness in candidate genes and neutral regions, with GCX being the most distinct from the rest in both types of markers, probably due to a strong genetic drift effect in this population as can be seen with the Ne value (Ne=5, Table 2.5).

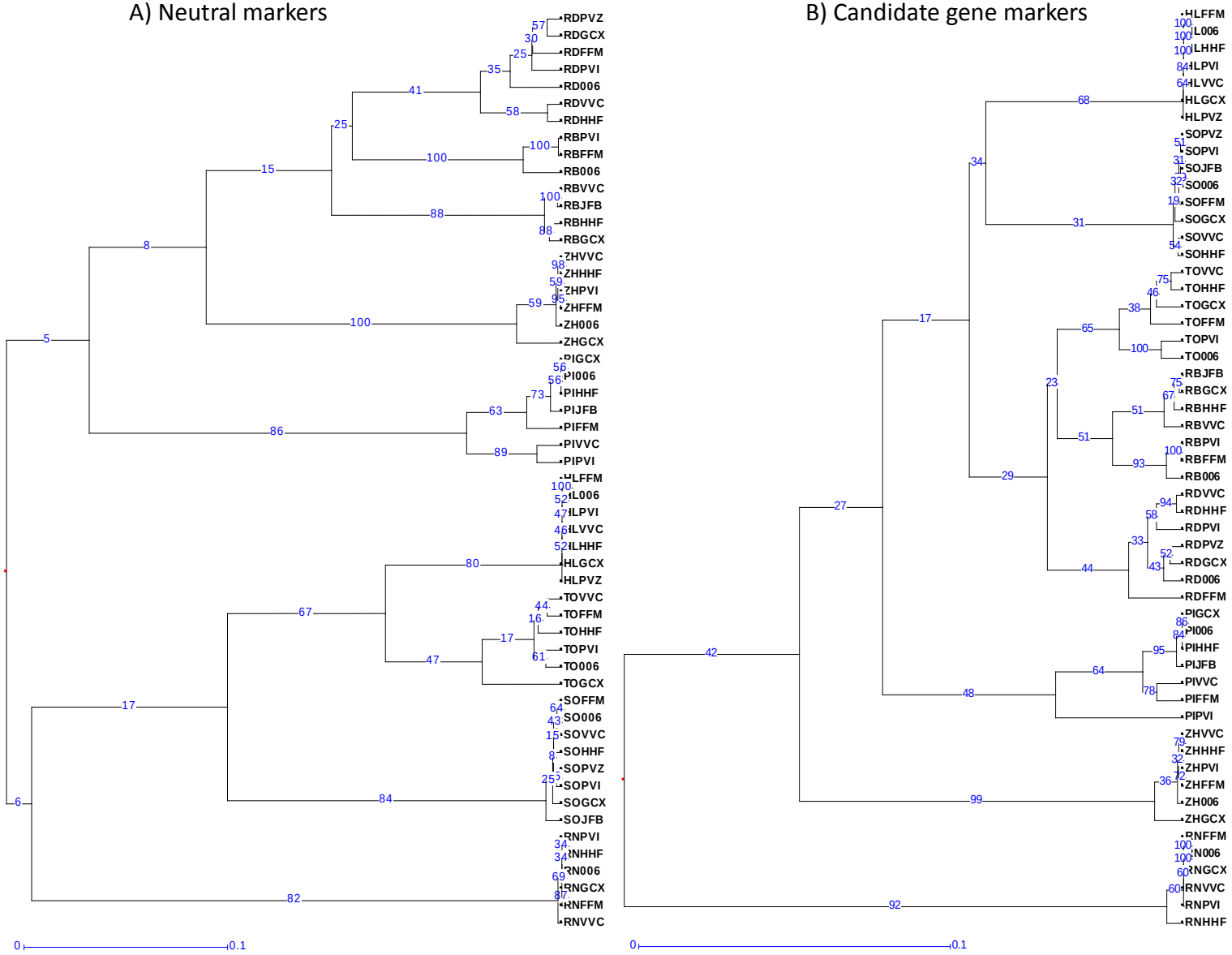


Figure 2.8: Phylogenetic trees of different versions of each variety A) for neutral markers and B) for candidate gene markers

2.3.4.3 Temporal evolution at the whole genome level: estimation of Ne

The genetic effective population sizes (N_e) were estimated on neutral markers (table 2.5) by comparing the allelic frequencies in each version of a variety to the frequencies in the initial population. All the versions for which Ne could not be estimated due to low Fc compared to sample size, were replaced by their demographic population size estimates based on the real number of plants grown and harvested for each variety in each farm. The genetic effective population size of different versions of a variety gives an indication of the level of temporal differentiation that occurred in each version. In Haute-Loire, for most of the versions, Ne was not estimable and we gave their demographic population size (FFM, HHF, PVI and VVC). This was due to the high level of allele fixation in the variety which did not allow to reveal variation in allele frequencies. Only GCX showed some differentiation. For Piave GCX and HHF, Ne was replaced by the demographic size indicating little temporal evolution in these populations, whereas the others showed drastically reduced Ne (values ranging from 3 to 19) showing strong genetic drift (and/or bottlenecks) in these populations. In Rouge de Bordeaux, all the versions except FFM appeared to be differentiated over time as they have *Ne* ranging from 3 to 22. In the FFM version, the low number of individuals used in the analysis prevented the detection of variation in allele frequencies (strong sampling effect compared to Fc) and also probably increased the possibility of removal of minor alleles causing lower Fc values. All versions of Redon appeared to be differentiated as *Ne* ranged from 6 to 22 (Table 2.5). As all Renan populations were highly fixed (Table 2.5), this led to little or no variation in allele frequencies (FFM, HFF and PVI) while VVC and to some extent GCX showed drastic variation at the few locus that were not fixed leading to small estimated Ne (5 and 33 respectively). Solina also did not show strong temporal shift as the lowest *Ne* was 49 for JFB version (Table 2.5) although the populations were highly diverse. This rather indicated that the population composition was rather stable over time and that no bottleneck or strong genetic drift occured. In Touselles, GCX and FFM appeared to be differentiated (*Ne* values 7 and 17 respectively). GCX in Zonne hoeve appeared divergent. Interestingly GCX versions of all the varieties except Piave showed strong temporal differentiation.

Table 2.5: Effective population size and sample size for 46 populations (2009 versions)

		FFM	GCX	HHF	JFB	PVI	PVZ	VVC
HL	$\mathbf{N_e}$	**6000**	**25**	**6000**	**-**	**10000**	**1063**	**6000**
	S_s	6	29	29	-	24	21	25
PI	$\mathbf{N_e}$	**8**	**4000**	**6000**	**19**	**3**	**-**	**5**
	S_s	28	28	23	25	17	-	23
RB	$\mathbf{N_e}$	**6000**	**3**	**3**	**4**	**22**	**-**	**4**
	S_s	11	29	27	30	25	-	25
RD	$\mathbf{N_e}$	**22**	**8**	**6**	**-**	**15**	**11**	**8**
	S_s	5	26	28	-	18	25	28
RN	$\mathbf{N_e}$	**191**	**33**	**6000**	**-**	**6000**	**-**	**5**
	S_s	29	28	23	-	25	-	29
SO	$\mathbf{N_e}$	**4000**	**97**	**2700**	**49**	**4000**	**81**	**6000**
	S_s	30	26	28	26	19	23	29
TO	$\mathbf{N_e}$	**17**	**7**	**40**	**-**	**121**	**-**	**6000**
	S_s	26	24	27	-	21	-	27
ZH	$\mathbf{N_e}$	**6000**	**5**	**6000**	**-**	**240**	**-**	**6000**
	S_s	30	29	23	-	23	-	27

Where $\mathbf{N_e}$ represents effective population size and S_s represent sample size

3.3.4.4 Testing for specific temporal differentiation at candidate genes markers

The significance of temporal differentiation at each candidate gene was tested by adapting a method developed by (Goldringer and Bataillon, 2004). The temporal evolution of the genetic group composition was also tested by using the same procedure, in order to assess whether some genetic groups might have a selective advantage in some populations. If so, it would not be possible to separate among specific effects of candidate genes or effect of the overall genetic background.

Overall, although 30 candidate gene markers among 34 showed at least one significant large change in frequency in one population of one variety, few cases of systematic significant temporal differentiation were observed (Table 2.6). Only 11 candidate gene markers (CA4, CA7, CA8, CA9, CA15, CA17, CA20, CA21, CA23, CA29, CA33; Table 2.6) showed significant

Table 2.6: Candidate gene markers showing significant temporal differentiation

Variety	Pop	DAPC Group significance	CA1	CA2	CA3	CA4	CA5	CA6	CA7	CA8	CA9	CA10	CA11	CA12	CA13	CA14	CA15	CA16	CA17	CA18	CA34	CA19	CA20	CA21	CA22	CA23	CA24	CA25	CA26	CA27	CA28	CA29	CA30	CA31	CA32	CA33
			PHYA	ZTL	VIL2	CO3	SMZ	Vrn1B	Vrn-1B	FT	SMZ	CO1	VIL2	PHYA	TaGI3	CO4	Vrn1B	CO1	LDD-A	TaHd1A	FTB	CO1	SMZ	SOC1	ZTL	CO4	ZTL	COAB	FT-A	PPD-A1	PPD-B1	PPD-B1	PPD-D1	TAGW2	VRN-1A	VRN-1D
HL	HLFFM	NA																																		
	HLGC1	NA																																		
	HLHHF	NA																																		
	HLPVI	NA																																		
	HLPVZ	NA							1															0.96						0.95						
	HLVVC	NA									1																									
	No. Of sign. CA		0		0	0	0	0	1	0	1	0	0		0	0	0	0	0	0	0	0	0	1		0	0	0	0	1		0	0	0	0	0
PI	PIFFM	0.955									1	1					1	0.98				0.98	0.99			0.99			1			0.97		0.99		0.95
	PIGCX	0.71																																		
	PIHHF	0.71	1				1	1		1							0.96						0.95			0.97										
	PIJFB	0.937																				1							1			0.99				
	PIPVI	0.965									1	1				1	0.97	0.97	0.97				0.96			0.99	1					0.99	1		0.98	0.98
	PIVVC	0.933					1				1					1			0.97				0.96			0.98	1					0.99	1			0.97
	No. Of sign. CA		1		0	0	2	1	0	1	3	2	0		0	2	3	2	2	0	0	2	4	0		4	2	0	2	0		4	2	1	1	3
RB	RBFFM																																			
	RBGCX	0.885							1										0.96																	
	RBHHF	0.812																												0.95						
	RBJFB	0.71							1																											
	RBPVI																																			
	RBVVC	0.655							1																											
	No. Of sign. CA		0		0	0	0	0	3	0	0	0	0		0	0	0	0	1	0	0	0	0	0		0	0	0	0	1		0	0	0	0	0
RD	RDFFM	1				1														1		1									0.96				
	RDGCX	0.978																						0.97												
	RDHHF	0.981								1																										
	RDPVI	0.999								1																										
	RDPVZ	0.966								1																										
	RDVVC	0.996								1																										
	No. Of sign. CA		0		0	1	0	0	0	4	0	0	0		0	0	0	0	0	1	0	1	0	1		0	0	0	0	0		1	0	0	0	0

The green color indicates the completely fixed markers. Purple color indicates the significantly differentiated DAPC groups

Table 2.6: continued

Variety	Pop	DAPC Group significance	CA1 PHYA	CA2 ZTL	CA3 VIL2	CA4 CO3	CA5 SMZ	CA6 Vrn1B	CA7 Vrn-1B	CA8 FT	CA9 SMZ	CA10 CO1	CA11 VIL2	CA12 PHYA	CA13 TaGI3	CA14 CO4	CA15 Vrn1B	CA16 CO1	CA17 LDD-A	CA18 TaHd1A	CA34 FTB	CA19 CO1	CA20 SMZ	CA21 SOC1	CA22 ZTL	CA23 CO4	CA24 ZTL	CA25 COAB	CA26 FT-A	CA27 PPD-A1	CA28 PPD-B1	CA29 PPD-B1	CA30 PPD-D1	CA31 TAGW2	CA32 VRN-1A	CA33 VRN-1D
RN	RNFFM	NA																																		
	RNGCX	NA																																		
	RNHHF	NA																															1			1
	RNPVI	NA																																		
	RNVVC	NA																																		
	No. Of sign. CA																																			
SO	SOFFM	0.711				1																					0.99								
	SOGCX	0.312							1																											
	SOJFB																																			
	SOPVI	0.963																					0.98													
	SOPVZ																																			
	SOHHF	0.899			1	1				1			1								1					0.99		1								
	SOVVC	0.406								1	1									1	1							0.98								
	No. Of sign. CA		0		1	2	0	0	1	2	1	0	1		0	0	0	0	0	1	2	0	0	1		1	0	3	0	0		0	0	0	0	0
TO	TOFFM	0.741				1				1									0.98														1			1
	TOGCX	0.833									1																									
	TOHHF	0.789				1											0.95						0.96	0.97		0.96				0.98						
	TOPVI	0.792				1																														
	TOVVC	0.753				1		1							1		0.97						0.97	0.97			1							0.97	0.97
	No. Of sign. CA		0		0	4	0	1	0	1	1	0	0		1	0	2	0	1	0	0	0	2	2		1	1	0	0	1		0	1	0	1	2
ZH	ZHFFM	0.419																										1								
	ZHGCX	0.584			1						1		1						0.98																0.98	
	ZHHHF																																			
	ZHPVI	0.426																										1								
	ZHVVC	0.846																						0.97												
	No. Of sign. CA		0		1	0	0	0	0	0	1	0	1		0	0	0	0	1	0	0	0	0	1		0	0	0	2	0		0	0	0	1	0

differentiation in more than four populations (maximum eight populations for CA8 corresponding to the *FT* gene), among which, 8 were candidate genes associated to flowering time in wheat. Different loci showed significant variation in specific varieties. Piave showed the highest number of genes with significant temporal differentiation as compared to other varieties, averaging seven loci per version showing significant temporal differentiation. Three loci, (CA20, CA23 and CA29) corresponding to the *SMZ*, *CO4* and *PPD-B1* genes showed significant variation in four PI versions, and three loci i.e. CA9, CA15 and CA33 (corresponding to *SMZ*, *VRN-1B* and *VRN-1D* genes respectively) showed significant variation in three versions of Piave (Table 2.6). Two Piave populations (PIFFM and PIPVI) have undergone a significant change in their genetic group composition, with a decrease of the frequency of group PI2 which was predominant in the initial population. Consistently, in each of the two populations, a higher number of loci showed significant variation (resp. 11 and 13), leading to a confusion of effects between the genetic background and the candidate genes variation.

CA7 (corresponding to *VRN-1B* gene) showed a significant change in frequency in three RB populations as compared to the initial RB population. Although the variation in genetic group frequency was significant in all RD populations, only one gene, CA8 (corresponding to *FT* gene), showed significant change in four RD versions compared to the initial RD population, while five additional loci significantly varied in a single population. This suggests that variation in group frequency was not driven by selection on flowering time or at least not on these candidate genes. Renan remained highly fixed except for HHF version in which two loci i.e. CA30 and CA33 (corresponding to *PPD-D1* and *VRN-1D* genes) showed significant differentiation from the initial RN population. In Solina, significant variation was observed at CA25 in three versions. Whereas one population (SOPVI) showed a significant change in the genetic group frequency, but it was not the one where the highest number of gene markers were detected (seven candidate gene markers in SOHFF). In Touselles, CA4 (corresponding to CO3 gene) revealed a significant change in four versions as compared to initial TO population. It was interesting to observe that in most varieties only one locus (different for different varieties) undergone differentiation for more than two versions.

2.3.5 Association of candidate genes with heading date and plant height

Analysis of variation test using model 1 (equation 2, where only the version effect was tested) revealed that heading date and plant height always showed significant differentiation (Table 2.7) among different versions (including the 2006 initial version) of each variety. This indicated that the evolution of these populations occurred in a very short time span even in the case of the modern variety Renan, although the range of variation among mean values of the versions strongly differed from one variety to the other with the variety Piave showing the largest differentiation (Table 2.8).

Table 2.7: ANOVA test for heading date and Plant height for each variety

Variety	Heading date			Plant Height		
	Model 1	Model 2		Model 1	Model 2	
	Pop	Pop	GG	Pop	Pop	GG
HL	***	NA	NA	***	NA	NA
PI	***	***	***	***	**	***
RB	***	***	***	*	*	***
RD	*	***	***	***	***	***
RN	**	NA	NA	***	NA	NA
SO	***	***	NS	**	**	NS
TO	*	***	***	**	***	***
ZH	***	***	***	*	**	***

+= Pvalue< 0.1, *= Pvalue< 0.05, **= Pvalue< 0.01, ***= Pvalue< 0.001,

The genetic groups (*GGj)* designated by DAPC (model 2; equation 3) also showed significant association with the heading date and plant height for all the varieties except Solina. Since Haute Loire and Renan had only one genetic group each, these two varieties were not tested with model 2. The means of genetic groups in each variety are given in Table 2.9.

The statistical association between selected candidate genes polymorphism (CA9, CA15, CA20, CA24, CA29 and CA33 for PI variety, CA7 for RB variety, CA8 for RD variety, CA25 for SO variety, CA4 for TO variety and CA26 for ZH variety; table 2.6), and heading date and plant height phenotypic variation was studied in two different samples in each of the varieties (except

Table 2.8: Heading date and Plant Height means for each version in each variety.

Heading date								
	HL	PI	RB	RD	RN	SO	TO	ZH
INI	1250 b	666.7 d	1055 c	1353 a	1194 a	1229 abc	1242 ab	1284 a
FFM	1236 b	804.9 abc	1100 abc	1355 a	1187 a	1227 bc	1252 ab	1296 a
GCX	1261 ab	765.2 bcd	1201 a	1302 ab	1184 ab	1215 bc	1216 ab	1206 b
HHF	1254 b	718 cd	1173 ab	1255 b	1179 ab	1249 ab	1173 b	1268 a
JFB	NA	754.1 bcd	1164 ab	NA	NA	1203 c	NA	NA
PVI	1242 b	901.1 ab	1120 abc	1319 a	1139 b	1208 bc	1280 ab	1301 a
PVZ	1285 a	NA	NA	1323 a	NA	1268 a	NA	NA
VVC	1266 ab	918.2 a	1100 bc	1329 a	1166 ab	1217 bc	1296 a	1294 a
Plant Height								
	HL	PI	RB	RD	RN	SO	TO	ZH
INI	119.8 b	70.34 c	120 ab	144.8 a	69.97 b	121.2 b	137.7 ab	97.73 a
FFM	116.2 b	75.18 c	120.5 ab	129.2 b	71.45 ab	119.1 b	126.4 b	91.32 b
GCX	130.7 a	77.46 bc	124.7 ab	126.6 b	69.96 b	123.1 ab	134.5 ab	94.21 ab
HHF	128.2 a	75.17 c	130.4 a	124.5 b	75.7 a	130.8 a	128.2 b	97.48 ab
JFB	NA	75.43 bc	124.5 ab	NA	NA	123.1 ab	NA	NA
PVI	123.4 ab	91.94 a	121.8 ab	136.5 ab	72.79 ab	125.7 ab	144.8 a	96.52 ab
PVZ	124.9 ab	NA	NA	131.8 b	NA	124.2 ab	NA	NA
VVC	126.6 ab	87.63 ab	117.8 b	135.7 ab	75.41 a	120.6 b	137 ab	98.38 a

Letters represent the level of significance among different versions in each variety

Table 2.9: Heading date and Plant Height means for different genetic groups in each variety.
A) Heading Date and B) Plant height

A) Heading date

PI		RB		RD		SO		TO		ZH	
Gene groupe	Heading date	Gene groupe	Heading date	Gene groupe	Heading date	Gene groupe	Heading date	Gene groupe	Heading date	Gene groupe	Heading date
HL	1274 a	RB2	1191 a	RD2	1357 a	SO2	1233 a	TO1	1349 a	ZH3	1293 a
PI1	916.1 b	RB1	1182 a	RD1	1356 a	SO1	1231 a	TO3	1314 ab	ZH2	1281 a
PI2	722.9 c	CG1	1013 b	RD3	1348 a	CG2	1229 a	HL	1264 abc	ZH1	951.3 b
				RD4	1311 ab	SO3	1228 a	TO2	1258 bc		
				HL	1254 b	SO4	1217 a	TO5	1249 bc		
				CG1	1133 c			CG2	1181 bc		
								TO4	1141 c		
								ZH1	696.7 d		

B) Plant height

Gene groupe	Plant height	Gene groupe	Plant height	Gene groupe	Plant height	Gene groupe	Plant height	Gene groupe	Plant height	Gene groupe	Plant height
HL	121.2 a	RB1	128.7 a	RD2	143.3 a	SO1	127.4 a	TO1	161 a	ZH2	98.36 a
PI1	79.48 b	RB2	126.3 a	RD1	141.8 a	SO4	123.4 a	TO3	144.5 b	ZH3	95.69 a
PI2	74.23 c	CG1	114.8 b	RD3	137.9 a	SO3	123.3 a	TO4	135.4 bc	ZH1	83 b
				RD4	125.1 b	SO2	122 a	TO5	130.8 c		
				HL	124.7 bc	CG2	119.6 a	HL	127.5 c		
				CG1	113.2 c			TO2	126.5 c		
								CG2	118.4 c		
								ZH1	84.33 d		

Letters represent the level of significance among different genetic groups in each variety

Table 2.10: Association between selected candidate genes and phenotypic traits. A) Heading date and B) PlantHeight

A) Heading date

Variety	Population	All versions						Versions with significant gene					
	ANOVA type	Type I			Type III			Type I			Type III		
	Gene	Gene	GG	Pop	Gene	GG	Pop	Gene	GG	Pop	Gene	GG	Pop
PI	CA9	***	***	+	*	***	+	***	***	NS	*	***	NS
	C15	NS	***	+	+	***	+	+	***	NS	NS	***	NS
	CA20	NS	***	+	NS	***	+	NS	***	NS	NS	***	NS
	CA23	***	***	NS	NS	***	NS	***	***	NS	NS	***	NS
	CA29	***	***	*	***	***	*	**	***	NS	NS	***	NS
	CA33	NS	***	NS	NS	***	NS	NS	***	NS	NS	***	NS
RB	CA7	***	***	NS	NS	**	NS	*	***	NS	***	**	NS
RD	CA8	*	***	*	NS	***	*	NS	***	NS	+	***	NS
SO	CA25	NS	NS	**	NS	NS	**	NS	NS	**	NS	NS	**
TO	CA4	*	***	+	NS	***	+	*	***	+	NS	***	+
ZH	CA26	*	***	*	NS	***	*	+	NS	**	NS	NS	**

B) Plant height

Variety	Population	All versions						Versions with significant gene					
	ANOVA type	Type I			Type III			Type I			Type III		
	Gene	Gene	GG	Pop	Gene	GG	Pop	Gene	GG	Pop	Gene	GG	Pop
PI	CA9	***	***	+	*	***	+	***	***	NS	*	***	NS
	C15	NS	***	+	+	***	+	+	***	NS	NS	***	NS
	CA20	NS	***	+	NS	***	+	NS	***	NS	NS	***	NS
	CA23	***	***	NS	NS	***	NS	***	***	NS	NS	***	NS
	CA29	***	***	*	***	***	*	**	***	NS	NS	***	NS
	CA33	NS	***	NS	NS	***	NS	NS	***	NS	NS	***	NS
RB	CA7	***	***	NS	NS	**	NS	*	***	NS	***	**	NS
RD	CA8	*	***	*	NS	***	*	NS	***	NS	+	***	NS
SO	CA25	NS	NS	**	NS	NS	**	NS	NS	**	NS	NS	**
TO	CA4	*	***	+	NS	***	+	*	***	+	NS	***	+
ZH	CA26	*	***	*	NS	***	*	+	NS	**	NS	NS	**

+= P value < 0.1, *= P value < 0.05, **= P value < 0.01, ***= P value < 0.001, NS = nonsignificant

Table 2.11: Heading date and Plant Height means for different candidate gene markers in different varieties. A) Heading date and B) Plant Height

A) Heading date

Heading Date	PI						RB	RD	SO	TO	ZH
	CA9	CA15	CA20	CA23	CA29	CA33	CA7	CA8	C25	CA4	CA26
AA	1026 a	899.3 a	899.5 a	1046 a	992.5 a	995 a	1190 a	1340 a	1242 a	1241 a	1312 a
BB	723.2 b	764 b	767.1 ab	736.4 b	713.7 c	775 b	1068 b	1268 b	1222 b	1239 a	1263 b
AB			994.7 a		833.7 b				1218 b		

B) Plant Height

Plant height	PI						RB	RD	SO	TO	ZH
	CA9	CA15	CA20	CA23	CA29	CA33	CA7	CA8	C25	CA4	CA26
AA	93.13 a	77.74 a	77.82 A	95.57 a	92.59 a	87 a	126.4 a	135 a	125 a	135 a	96.8 a
BB	74.07 b	77.12 a	77.46 A	74.89 b	73.21 b	77.6 a	119.1 b	129.9 b	123 a	128 b	93.7 b
AB			78 A		84.67 a				124 a		

Letters represent the level of significance among different homozygote and heterozygote for each gene in each variety. Due to biallelic nature of genotyping, AA and BB represent homozygote genotypes whereas AB represents heterozygote genotype

for HL and RN). Sample-I: all versions and Sample-II: versions with significant differentiation at the studied gene. For each level, ANOVA type I and type III were used.

All the studied candidate genes (except CA4 in Touselles) showed significant effect on heading date (Table 2.10A) when gene effect was given maximum preference (i.e. not adjusted for the other effects, ANOVA type I) in both samples (except CA33 which was non-significant at sample-II) although the level of significance varied among different genes. Interestingly, when the gene effect was tested with minimum preference (adjusted for population and group effects, ANOVA type III), only four genes (CA9 and CA29 in Piave, CA8 in Redon and CA26 in Zonne Hoeve) remained significant in sample I, whereas three genes, CA7 in Rouge de Bordeaux, CA25 in Solina and CA26 in Zonne Hoeve showed significant effect in sample II. These results indicate a certain level of association between the genes showing significant temporal differentiation (table 2.10A) at multiple versions in a variety and the heading date. The difference between type I and type III tests suggests that the polymorphism at these genes is correlated tothe genetic group and/or to the population which does not excludes the possibility that a candidate gene is indeed associated to the phenotypic trait and submitted to selection in these populations but prevents the detection.

For plant height, comparatively low level of association between genes and plant height was observed (table 2.10B). CA9, CA23 and CA29 in Piave, CA7 in RB, CA8 of Redon, CA4 in Touselles, and CA26 in Zonne hoeve showed significant association with plant height when ANOVA type I (with maximum preference to genes) was used with sample I as well as sample II (with the exception of CA15 in Piave). Almost all the genes showed no association (except CA9, CA15 and CA29 in Piave) when ANOVA type III was used with sample I and with sample II (except CA9 and CA15 in Piave). These results indicate that low level of association between candidate genes markers showing significant temporal differentiation (table 2.11) at multiple versions in a variety and the plant height. The means of heading date and plant height for candidate genes in different varieties are given in table 2.11.

2.4 Discussion

2.4.1 Fine genetic structure of farmers' varieties and influence of conservation method

Since the start of domestication around ten thousand years ago (Diamond, 2002), humans have influenced the population genetic structure of the plant and animal species, both intentionally and unintentionally (Chapin Iii *et al.*, 2000). More recently in the 20th century, when crop diversity maintenance became a concern, the conservation methods used to safeguard the genetic resources such as the landraces and historic varieties, have influenced their genetic structure and within variety diversity (Soleri and Smith, 1995; Tin *et al.*, 2001; Gómez *et al.*, 2005; Rice *et al.*, 2006; Negri and Tiranti, 2010). The farmers' varieties used in this study included a diversified set of landraces, mixtures and historic varieties, ranging from landraces, such as Piavé and Haute loire, which were conserved ex situ and recently obtained by the farmers from the genebank to the landrace Solina which was always conserved in situ by farmers, and from mixtures of landraces such as Redon and Touselles, which were constructed by farmers (after obtaining samples of the components from gene banks) and have been conserved *in situ* for the last 10 years to the historical variety Rouge de Bordeaux which has been always conserved *in situ* by farmers. In addition, we studied a mixture of two modern varieties (Zonne Hoeve) made by a farmer and grown for 15 years on farm and a modern variety, Renan which was used as a reference for a classical modern variety. Therefore these farmers' varieties provide an opportunity to study their genetic structure in order to compare the influence of two conservation methods (*ex situ* and *in situ*) on the "within diversity" of these farmers' varieties and to analyse the impact of the classical farmers' practice of mixing varieties and resowing the mix over time. In first part of this study, we investigated the genetic structure of farmers' varieties (landraces, historic varieties and mixtures) and how the history of these varieties has been influencing them.

Out of the three landraces studied, Haute loire and Piave were conserved *ex situ* whereas Solina was always conserved *in situ*, and continuously cultivated in its area of origin by the farmers. Haute loire, was found to be highly homogeneous with a single genetic group and very low within-variety genetic diversity, which appeared similar to diversity within the modern variety Renan. Although landraces are supposedly diverse, the observed genetic structure of Haute loire is consistent with a genetic bottleneck that might have happened eitheir when the sample was

collected or during the regeneration process as reported in Parzies *et al.*, (2000). It is more likely that the bottleneck occured during *ex situ* collect or conservation than during the time spent under on-farm cultivation (only one year after the farmer obtained it from the gene bank) because it was grown in a 20 m^2 plot that would harbour a minimum of 4000 plants. Therefore this highly fixed version of Haute loire suggests that the total ancestral diversity of the landrace was not conserved *ex situ* as the sample used for the conservation purpose could be a very small part of the initial diversity. The other landrace conserved *ex situ*, Piavé, showed one prevailing group consisting of a main haplotype and a few connected less frequent haplotypes and another smaller distinct group as revealed by haplotypic analysis (figure 2.4). Thus, this indicated that the process of *ex situ* conservation has maintained more diversity than in the case of Haute-Loire and that either two different sources were used or the ancestral Piavé landrace was constituted by two or more genetic lines which evolved all together but kept their individuality. Such composite structure (that can also be described as polyclonality) has been found for the *in situ* conserved landrace Solina as well as for the historic variety Rouge de Bordeaux in this study and in a previous study comparing Rouge de Bordeaux samples from *in situ* and *ex situ* conservation (Thomas *et al.*, 2012).

Thus, the third landrace Solina, which was under continuous cultivation in its area of origin and was never conserved *ex situ* revealed a highly diverse and complex genetic structure, divided in four genetic groups (DAPC analysis; figure 2.2), with a very high number of haplotypes most of which being very close to each other. This complex genetic structure of Solina might be the result of a very long history in the territory of Abruzzo, as witnessed by historical documents such as bills of sale from the tradefair of Lanciano in 1500 and in a text of Michele Torcia from the 18th century (http://www.abruzzoeappennino.com/magazine_articoli.asp?id=311 and in French: http://www.spicilege.org/index.php?option=com_adsmanager&page=show_ad&adid=170&catid=12&Itemid=0). The comparison with Haute loire and Piave provides insights about the impact of the two conservation methods in use for the protection of genetic diversity. Apart from providing evidence that the on-farm conservation allows greater level of within diversity as compared to *ex situ* conservation, it also confirms that the genetic structure of these three landraces, which shows a gradient in diversity and complexicity from Haute-Loire to Piavé and to Solina, is in accordance with their historic perspective.

Rouge de Bordeaux, the historic variety used in this experiment, also showed a high level of within variety diversity. Three groups were detected by the structure analysis (two specific main groups and one group shared with Redon). The haplotypic analysis revealed two groups mainly consisting in a highly frequent haplotype and a third one that was distributed other several connected haplotypes, with all these three groups being very distinct from each other. These results are compatible with the previously reported genetic structure of this historic variety (Thomas *et al.*, 2012), where four main haplotypic groups had been detected but only three were present in the specific population of origin in this study, the JFB population (sample of of 2003, JFB03 and sample of 2006, JFB06). Yet, it is not easy to make the connection since the most frequent group in (Thomas *et al.*, 2012, green in figure 5) was also the most diverse, which is not the case here. It may be due to some bottleneck effect at the starting of the project. The difference between both studies might also stem from the differences in the markers used since the previous study on the Rouge de Bordeaux populations was based on SSR markers which have a higher evolutionary rate than the SNP used in this study. This nonetheless complexe and diverse structure of Rouge de Bordeaux further strengthens the argument of the effectiveness of on-farm conservation method to maintain within-variety diversity, since this historic variety has always been conserved *in situ*.

All these results showed that the genetic structure of these farmers' varieties and landraces are rather consistent with their history of origin. Interestingly the conservation method used for these farmers' varieties highly influenced their within variety diversity and the population structure. The landraces which were conserved *ex situ* (Haute loire and Piave) before they were obtained from the gene banks by the farmers and conserved *in situ* show lower diversity compared to the landrace (Solina) and historical variety (Rouge de Bordeaux) which were always conserved *in situ*. Few studies have documented the comparison of the effects of *ex situ* and *in situ* conservation on crop populations genetic diversity since most often, diversity of genetic resources is studied by analysing one individual or a mix of five or less individuals, that represent the landrace or accession, thus neglecting the within-variety (within-accession) diversity level (e.g. Roussel et al. 2004 & 2005). Sun *et al.*, (2012) found a higher genetic diversity and more alleles in landraces conserved *in situ* compared to the landraces conserved *ex situ* in rice. The decrease in genetic diversity due of genetic drift during regeneration with limited sample size in *ex situ* conservation has been shown in bean (Gómez et al. 2005) and barley (Parzies *et al.*, 2000).

While 88% of total natural diversity was found in *ex situ* conserved *Vatica guangxiensis* (Li *et al.*, 2002), a better maintenance of genetic diversity of *Parashorea chinensis* was observed in *in situ* conservation than in *ex situ* conservation (Li *et al.*, 2005). The effects of conservation of crop biodiversity by the two methods (*in situ* and *ex situ*) was studied in more details by Negri and Tiranti (2010) who identified reduced population size as the main factor causing the change in the genetic variation followed by *ex situ* multiplication which increased the subpopulation differentiation due to different environmental conditions than that of area of adaptation. On the other hand, the evolutionary mechanisms that allow for the maintenance and development of within-variety genetic diversity have been described more often in studies of traditional agricultural systems. It has been shown that gene flows (through pollen for outcrossing species and seeds for all species) was a key factor to within-variety diversification (e.g. on maize: (Louette *et al.*, 1997; Louette and Smale, 2000), on barley (for a review see Thomas et al. 2011). In these studies, diversity was found within each population cultivated under the name of a given landrace variety and to a lesser extend among those populations. Such heterogeneity is not incompatible with good productivity and quality under low input conditions probably due to good local adaptation and more buffering capacity under stressing conditions, as shown for instance in the case of an Italian landrace of celery (Torricelli *et al.*, 2013).

Since both methods of conservation (*ex situ* and *in situ*) have their advantages and disadvantages, they should not be viewed as alternatives of one another. In the present situation where a large amount of landraces are conserved *ex situ* (Plucknett, 1987), using mixtures to increase the within variability of the *ex situ* conservated landraces, could be a good approach. Therefore two mixtures of landraces and one mixture of modern varieties were also included in this experiment.

The two mixtures of landraces used in this experiment, Redon and Touselles revealed the high levels of genetic diversity within the population. Redon revealed 4 distinct genetic groups along with a shared group with Rouge de Bordeaux (DAPC analysis; figure 2.2). The haplotypic analysis revealed a high number of haplotypes, most of them being different from each other (Figure 2.4). This pattern of relatively higher number of quite distinct haplotypes of this mixture is in accordance with its history because this mixture was constructed by mixing 7 different sources or accessions. Similar pattern of genetic structure was revealed by Touselles. Interestingly, the level of relatedness among “Touselles” haplotypes was higher than that of

Redon haplotypes. This was not expected as the Touselles was constructed by mixing three *T. aestivum* landraces and one *T. turgidum* L. landrace. Another mixture "Zonnehoeve", which was created by mixing two modern German varieties but is continuously under cultivation for more than 10 years, was also used in this experiment. The structure analysis revealed two main specific groups that likely correspond to the two varieties along with a small additional group which also had some individuals in Touselles and in Piavé. Haplotypic analysis revealed few intermediate haplotypes that suggest recombination among the two initial varieties as well as mutation or gene flows during the 10 years of *in situ* conservation.

The results of these mixtures suggest elements towards a model to improve the within diversity of the accessions conserved *ex situ* and reduce the risk of failure due to stochastic events while reintroducing these populations. However while creating these kinds of mixtures, there could be risk of outbreeding therefore using the accessions which originated from similar environmental conditions could reduce the risk of outbreeding depression (Maschinski *et al.*, 2013) is recommended.

2.4.2 *Spatio-temporal* differentiation

Environmental variability with varying and uneven patterns of temperature and precipitations have been predicted for near future by scientists (Olesen *et al.*, 2011). Most of the high yielding, elite modern varieties give high yields across a wide range of geographic areas but within a narrow range of production conditions and survive severe stress through the heavy use of costly inputs (Ceccarelli, 1996) which rely on fossil fuels for their production. In this scenario, using heterogeneous populations (genetically diverse) with an ability to buffer the biotic and abiotic stresses by the virtue of their within-variety diversity provides a way towards a more sustainable agriculture (Wolfe *et al.*, 2008; Ostergaard *et al.*, 2009). The study of the *spatio-temporal* differentiation of different types of variety (farmers' varieties and a modern registered variety) after they were transplanted in very contrasted agro-climatic conditions in The Netherland, Italy and France provided elements to understand the underlying short term evolutionary processes.

Spacio-temporal differentiation among different versions (of these varieties) was detected at both genetic and phenotypic level, although these varieties were cultivated in the contrasting agro-climatic conditions for only three years. As shown in the previous section, the farmers' varieties

Table 2.12: Diversity indexes computed for all 54 populations based on 41 neutral markers

Population	H_e	H_o	Population	H_e	H_o
HL006	0	0	RDVVC	0.3067	0
HLFFM	0	0	RN006	0.0052	0.0039
HLGCX	0.0018	0	RNFFM	0.0008	0.0008
HLHHF	0.0008	0.0008	RNGCX	0	0
HLPVI	0	0	RNHHF	0.004	0.0042
HLPVZ	0.0057	0.0012	RNPVI	0.0037	0.0039
HLVVC	0	0	RNVVC	0.0124	0.0235
PI006	0.0391	0.0021	SO006	0.1714	0.0068
PIFFM	0.1991	0.0009	SOFFM	0.1455	0.0024
PIGCX	0.0416	0	SOGCA	0.1473	0.0113
PIHHF	0.0456	0.0011	SOGCB	0.1432	0.0205
PIJFB	0.1268	0.0098	SOHHF	0.1786	0.0009
PIPVI	0.3528	0.0029	SOJFB	0.1498	0.0216
PIVVC	0.2851	0.0064	SOPVZ	0.1475	0.0127
RB006	0.3204	0.0028	SOVVC	0.1519	0.0042
RBFFM	0.3852	0	TO006	0.3144	0.0081
RBGCX	0.1968	0	TOFFM	0.231	0.0056
RBHHF	0.2727	0.0081	TOGCX	0.2746	0.001
RBJFB	0.2967	0	TOHHF	0.332	0
RBPVI	0.3731	0.001	TOPVI	0.2711	0.007
RBVVC	0.2927	0.0068	TOVVC	0.2816	0.0047
RD006	0.2272	0.0084	ZH006	0.1395	0.0055
RDFFM	0.1821	0	ZHFFM	0.1355	0.0008
RDGCX	0.2687	0.001	ZHGCA	0.2462	0.0025
RDHHF	0.3142	0	ZHGCB	0.1186	0.0021
RDPVI	0.234	0.0163	ZHHHF	0.1518	0.0042
RDPVZ	0.2979	0.001	ZHVVC	0.1233	0.0036

With H_e: unbiased Nei's estimate of genetic diversity (Nei 1978), H_o: mean observed heterozygosity

used in this experiment demonstrated varying level of within-variety diversity and structure, therefore each of these varieties responded to the environmental variation differently. Factorial Analysis of Correspondences (AFC) was used to study the overall differentiation within a variety and among different varieties. AFC analysis provided a first overview of the genetic differentiation among different versions of each variety in space and time. It revealed that varieties like Haute loire and Renan, with very low within-variety diversity (*He* values 0 and 0.005 respectively; table 2.12) showed low level of differentiation. In the case of Renan, it is interesting to note that although most versions were highly fixed (with He less than 1%), consisting only in one main haplotype and one which was connected, at the neutral markers level, a significant differentiation was detected at the phenotypic level for both plant height and heading date with some versions becoming later (~ 3 days for PVI) and taller (~ 5 cm for HHF and VVC) than to the initial population. It is noteworthy that two genes were detected as significantly varying in HHF compared to the initial population. Testing for their association with both phenotypic traits will be the next step.

Similar behavour was observed for Haute Loire in which a significant differentiation was detected at the phenotypic level for both plant height and heading date with some versions becoming later (~ 3 days for PVZ) and taller (~ 11 cm for HHF and VVC) than the initial population. Four genes were detected as significantly varying, 3 of them showed this variation in PVZ. It would be interesting to test their association with both traits.

Interestingly, significant phenotypic differentiation among highly fixed varieties like Renan and landrace Haute loire was detected. Although Renan is a modern variety and thus highly fixed to correspond to the criteria Distinction Uniformity Stability of the registration, there might be due to some residual variation at genes controlling the phenotypic traits. In the case of Haute loire, variation at candidate genes and possibibly at other unstudied genetic regions could cause this phenotypic response. There is also a possibility of involvement of epigenetic variation as they don’t involve any change in the DNA sequence.

Piave showed low within-variety diversity in the initial 2006 population while more diversity was found within the 3rd year versions and they seemed to have undergone relatively strong differentiation. The low diversity within the initial population could be due to the sampling effect (since only a small proportion of the initial seed lot was analysed, sample size was 28). This

supposition is strengthened by the fact that the PI1 genetic group (detected by DAPC analysis in the previous section) which was found in other Piave versions was absent from the initial population. As most of the versions appeared to be diverse (He ranging from 0.126 to 0.39) except GCX and HHF, Piave showed higher level of differentiation with AFC analysis and the phylogenetic tree. With the exception of two fixed versions (GCX and HHF), the low Ne (ranging from 3 to 19) confirmed the higher level of differentiation. The phenotypic evolution was rather unidirectional as the heading date tended to become later in the 2009 versions compares to the 2006 initial version (~ 10 days for PVI and VVC). Plant height also increased in all populations (~ 14 cm in PVI) even for GCX (farm of origin). This higher variation in PVI and VVC versions, to some extent, can correspond to the presence of HL1 group in these versions as HL1 revealed to be much later (~ 10 days) and taller (~ 5 cm) (Table 2.8). A large number of candidate gene markers (as many as 20) showed significant variation in at least one version as compared to the initial population. Although only selected candidate gene markers (6 markers) were tested for association with phenotypic traits, all the tested candidate gene markers showed certain level of association with the heading date and only three markers were found to be associated with plant height (Table 2.10). It is not clear based on these results is these HL1 individuals might not have arrived as migrants in PVI and VVC during seed manipulation and this needs to be further investigated.

Redon and Touselles are both mixtures and had high within-version diversity in the initial 2006 population (respective H_e values 0.227 and 0.314, Table 2.12). Both of these mixtures showed higher level of *spatio-temporal* differentiation as shown by AFC and phylogenetic tree. Redon appeared to be more differentiated than Touselles as all Redon versions showed low *Ne* whereas only FFM and GCX showed low *Ne* in Touselles (Table 2.12). This is consistant with the finding (in the previous section) that Touselles versions used in this study are slightly less diverse than reported by (Thomas, 2011). Both mixtures showed significant phenotypic differentiation (on both heading date and plant height) but not very strongly (5% level of significance for heading date and 1% for plant height, Table 2.8). This could be due to the higher buffering ability to contrasting agro-climatic conditions due to high within-version genetic diversity. Interestingly only HHF version appeared to show strong decrease in heading date in Redon showing relatively strong phenotypic differentiation that could be due to different climatic conditions of southern France (high temperature in early summer, Dawson *et al.*, 2013). A relatively larger number of

genes showed significant shift from initial 2006 population in Touselles as compared to Redon where only one candidate gene marker showed significant variation in four different versions (in addition to 5 markers showing variation in only one of the versions). This marker showed stronger association with heading date than with plant height in Redon. In Touselle, the tested candidate gene marker was not associated to heading date though it showed a certain level of association with plant height. This argues for the need of testing other candidate gene markers for association.

Zonne Hoeve is also a mixture but of two modern varieties. It had low within-version diversity in the initial 2006 population (H_e = 0.139) It showed low differentiation (shown by AFC, phylogenetic tree, Ne, heading date analysis). An exception is GCX version as it can be seen from AFC, phylogenetic tree and low *Ne* value. This was due to some individuals in GCX which constitute ZH1 genetic group. This is a much earlier and shorter group which has individuals mainly in GCX but also in HHF (one individual) and, one individual from VVC, and also shares some individuals in Touselles variety although Touselles individuals are different at some loci. One hypothesis is that this group was one of the two varieties (Bussard or Rektor) initially included in the mixture which had been conter-selected in the mixture due to its short size and unadapted earliness during the 15 years of cultivation on farm. Then, this group might have been more adapted in GCX and HHF (Southern sites). The other hypothesis is to consider the group ZH1 as migrants. Bussard has been released in 1963 and Rektor in 1980 and thus might differe in phenotypes, but more information on the two varieties is needed to really determine the most likely hypothesis.

Rouge de Bordeaux is a historic landrace with high within-version diversity in initial 2006 population. It showed high level of differentiation among different versions as shown by AFC. Phylogenetic tree revealed distribution of the different versions into two groups. Interestingly, in the phylogenetic tree with neutral markers, one group was specific and the other group was clustered with Redon. This might be due to the CG1 genetic group shared between Redon and Rouge de Bordeaux. The high temporal differentiation was evident from *Ne* values of different versions (ranging from 3 to 22 except for FFM, Table 2.5). Phenotypic differentiation at both phenotypic trait was observed, where the versions tend to be later (~ 10 days for GCX) and taller (~10 cm for HHF) from the initial 2006 population (except VVC for plant height). Structural

analysis in the previous section revealed that although, the number of haplotypes of Rouge de Bordeaux was low, these haplotypes were very distinct from each other. This could have a role in differentiation. Only three gene markers showed significant variation from initial 2006 population, one of which showed variation in three of the versions. This gene showed strong association with both phenotypic traits.

Solina, a landrace which was continuously conserved *in situ*, showed a low level of differentiation (based on AFC, phylogenetic tree and *Ne* value). Although it showed significant phenotypic differentiation at both traits, the means for heading date did not to vary strongly (the largest difference was three days and 11 cm at maximum, Table 2.8). Structural analysis in the previous section revealed a huge number of haplotypes but they were highly connected to each other. This could support the idea that this diversity provided a buffering effect and stabilized the phenotype.

In case of Solina and Rouge de Bordeaux, it appeared that if the haplotypes are more connected to each others, although high in number (as in Solina) it might provide buffering effect against the contrasting climatic conditions, whereas if the haplotypes are highly distinct (as in Rouge de Bordeaux), it might lead to rapid differentiation.

With these analyses, we also could see that the methods designed to study genetic differentiation among populations (AFC, AMOVA, phylogenetic trees) did not managed to give a complete vision of the whole structure of the varieties. It is probably due to the fact that the different varieties have different level of within-population genetic diversity and structure, which affect the paremeters used to estimate divergence among populations such as *Fst*. Therefore, the finer insight given by the DAPC method that allows to detect genetic groups specific or common to the different varieties and by the haplotypic network completed the picture efficiently. So, we obtained a better understanding of the genetic and phenotypic responses of all varieties to new environmental conditions.

2.5 Conclusions

The conservation history of these farmers' varieties influenced their genetic structure. *Ex situ* conservation seems to decrease within-variety genetic diversity whereas *in situ* conservation i.e. on-farm conservation tends to maintain and create within-variety genetic diversity in a dynamic and continuously evolving manner. Thus, to take advantage of both methods, one suggestion might be to reintroduce the *ex situ* conserved variety on farm, using mixture of relative landraces as it could increase the genetic diversity.

In response to cultivating these farmers' varieties in contrasting environmental conditions, certain level of *spacio-temporal* differentiation was observed. Different candidate gene loci showed significant temporal differentiation in different versions of varieties. Candidate genes showed association with heading date and plant height. The significant phenotypic differentiation in highly fixed farmers variety and modern variety might indicate the presence of variation at epigenetic level. It would be interesting to explore this avenue.

References

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, *et al.* 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics **11**, 702.

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.

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.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **115**, 721–733.

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**.

Bellon MR. 1997. Understanding farmer seed and variety systems On-farm conservation as a process: an analysis of its components. In: USING DIVERSITY Enhancing and Maintaining Genetic Resources On-farm. Available from http://www.idrc.ca/en/ev-85109-201-1-DO_TOPIC.html . Accessed Jan 2011

Bent AF. 2003. Crop diseases and stratigies for their control. In: Chrispeels MJ, Sadava DE. eds. *Plants, Genes, and Crop Biotechnology*. Jones & Bartlett Learning, 390-413.

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.

Vom Brocke K, Christinck A, Weltzien RE, Presterl T, Geiger HH. 2003. Farmers' Seed Systems and Management Practices Determine Pearl Millet Genetic Diversity Patterns in Semiarid Regions of India. Crop Science **43**, 1680–1689.

Vom Brocke K, Presterl T, Christinck A, Weltzien E, Geiger RHH. 2002. Farmers' seed management practices open up new base populations for pearl millet breeding in a semi-arid zone of India. Plant Breeding **121**, 36–42.

Callow JA, Ford-Lloyd B, Newbury HJ. 1997. *Biotechnology and plant genetic resources: conservation and use*. CAB International.

Ceccarelli S. 1996. Adaptation to low high input cultivation. Euphytica **92**, 203–214.

Chao S, Dubcovsky J, Dvorak J, Luo M-C, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S, *et al.* 2010. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.). BMC Genomics **11**, 727.

Chapin Iii FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, Sala OE, Hobbie SE, *et al.* 2000. Consequences of changing biodiversity. Nature **405**, 234–242.

Le Clerc V, Bazante F, Baril C, Guiard J, Zhang D. 2005. Assessing temporal changes in genetic diversity of maize varieties using microsatellite markers. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **110**, 294–302.

Dawson JC, Goldringer I. 2012. Breeding for genetically diverse populations: variety mixtures and evolutionary populations. Organic Crop Breeding, 77–98.

Dawson JC, Serpolay E, Giuliano S, Schermann N, Galic N, Berthellot J-F, Chesneau V, Ferté H, Mercier F, Osman A, *et al.* 2013. Phenotypic diversity and evolution of farmer varieties of bread wheat on organic farms in Europe. Genetic Resources and Crop Evolution **60**, 145–163.

Dawson JC, Serpolay E, Giuliano S, Schermann N, Galic N, Chable V, Goldringer I. 2012. Multi-trait evolution of farmer varieties of bread wheat after cultivation in contrasting organic farming systems in Europe. Genetica **140**, 1–17.

Elias M, Penet L, Vindry P, McKey D, Panaud O, Robert T. 2001. Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (Manihot esculenta Crantz), in a traditional farming system. Molecular ecology **10**, 1895–1907.

Engels J. 2002. Managing plant genetic diversity [electronic resource]. CABI.

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.

European Commission. 2008. Commission Directive 2008/62/. EC of 20 June 2008 providing for certain derogations for acceptance of agricultural landraces and varieties which are naturally adapted to the local and regional conditions and threatened by genetic erosion and for marketing of seed and seed potatoes of those landraces and varieties (Text with EEA relevance). Availableonline http://eur-lex.europa.

Finckh MR. 2008. Integration of breeding and technology into diversification strategies for disease control in modern agriculture. European Journal of Plant Pathology **121**, 399–409.

Fowler C, Hawtin GC, Hodgkin T. 2000. Foreword. Genes in the Field: On-farm Conservation of Crop Diversity. Rome, Italy: Bioversity International, .

Garrick RC, Sunnucks P, Dyer RJ. 2010. Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. BMC Evolutionary Biology **10**, 118.

Goldringer I, Bataillon T. 2004. On the distribution of temporal variations in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection. Genetics **168**, 563.

Gómez OJ, Blair MW, Frankow-Lindberg BE, Gullberg U. 2005. Comparative Study of Common Bean (Phaseolus vulgarisL.) Landraces Conserved ex situ in Genebanks and in situ by Farmers. Genetic Resources and Crop Evolution **52**, 371–380.

Hajjar R, Jarvis DI, Gemmill-Herren B. 2008. The utility of crop genetic diversity in maintaining ecosystem services. Agriculture, Ecosystems & Environment **123**, 261–270.

Henry JP, Pontis C, David J, Gouyon PH. 1991. An experiment on dynamic conservation of genetic resources with metapopulations. Species conservation: a population-biological approach. A. Seitz, V. Loeschcke, 185–198.

Jarvis DI, Brown AHD, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, Gyawali S, Tanto T, Sawadogo M, Mar I, Sadiki M, *et al.* 2008. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. Proceedings of the National Academy of Sciences **105**, 5326–5331.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**, 94.

Li Q, He T, Xu Z. 2005. Genetic evaluation of the efficacy of in situ and ex situ conservation of Parashorea chinensis (Dipterocarpaceae) in Southwestern China. Biochemical genetics **43**, 387–406.

Li Q, Xu Z, He T. 2002. Ex situ genetic conservation of endangered Vatica guangxiensis (Dipterocarpaceae) in China. Biological Conservation **106**, 151–156.

Louette D, Charrier A, Berthaud J. 1997. In Situ conservation of maize in Mexico: Genetic diversity and Maize seed management in a traditional community. Economic Botany **51**, 20–38.

Louette D, Smale M. 2000. Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica **113**, 25–41.

Maccaferri M, Sanguineti MC, Donini P, Tuberosa R. 2003. Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **107**, 783–797.

Maschinski J, Wright SJ, Koptur S, Pinto-Torres EC. 2013. When is local the best paradigm? Breeding history influences conservation reintroduction survival and population trajectories in times of extreme climate events. Biological Conservation **159**, 277–284.

Morin SR, Calibo M, Garcia-Belen M, Pham JL, Palis F. 2002. Natural hazards and genetic diversity in rice. Agriculture and Human Values **19**, 133–149.

Negri V, Tiranti B. 2010. Effectiveness of in situ and ex situ conservation of crop diversity. What a Phaseolus vulgaris L. landrace case study can tell us. Genetica **138**, 985–998.

Almekinders CJM, Elings A. 2001. Collaboration of farmers and breeders: Participatory crop improvement in perspective. Euphytica **122**, 425–438.

Altieri M, Anderson MK, Merrick L. 1987. Peasant Agriculture and the Conservation of Crop and Wild Plant Resources. Conservation Biology **1**, 49–58.

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.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **115**, 721–733.

Bellon MR. 1997. Understanding farmer seed and variety systems On-farm conservation as a process: an analysis of its components. USING DIVERSITY Enhancing and Maintaining Genetic Resources On-farm.

Bent AF. 2003. Crop diseases and stratigies for their control. Plants, Genes, and Crop Biotechnology. Jones & Bartlett Learning, .

Berger SL. 2007. The complex language of chromatin regulation during transcription. Nature **447**, 407–412.

Berthaud J, Clément JC, Emperaire L, Louette D, Pinton F, Sanou J, Second G. 2001. The role of local-level geneflow in enhancing and maintaining genetic diversity. Broadening the genetic base of crop production. CABI Publishing, 452.

Bird A. 2002. DNA Methylation Patterns and Epigenetic Memory. Genes & Development **16**, 6–21.

Blake NK, Lehfeldt BR, Lavin M, Talbert LE. 1999. Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. Genome / National Research Council Canada = Génome / Conseil national de recherches Canada **42**, 351–360.

Bossdorf O, Arcurri D, Richards C, Pigliucci M. 2010. Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology, 541–553.

Bossdorf O, Richards CL, Pigliucci M. 2008. Epigenetics for ecologists. Ecology letters **11**, 106–115.

Bost B, de Vienne D, Hospital F, Moreau L, Dillmann C. 2001. Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. Genetics **157**, 1773–1787.

Brown AHD, Hodgkin T. 2007. Measuring, managing and maintaining crop genetic diversity on farm. Managing biodiversity in agricultural ecosystems, 13–33.

Camacho Villa T, Maxted N, Scholten M, Ford-Lloyd B. 2005. Defining and identifying crop landraces. Plant genetic resources: characterization and utilization **3**, 373–384.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, *et al.* 2012. Biodiversity loss and its impact on humanity. Nature **486**, 59–67.

Ceccarelli S. 1994. Specific adaptation and breeding for marginal conditions. Euphytica **77**, 205–219.

Ceccarelli S. 1996. Adaptation to low high input cultivation. Euphytica **92**, 203–214.

Ceccarelli S. 2009. Evolution, plant breeding and biodiversity. Journal of Agriculture and Environment for International Development **103**, 131–145.

Cervera MT, Ruiz-García L, Martínez-Zapater JM. 2002. Analysis of DNA methylation in Arabidopsis thaliana based on methylation-sensitive AFLP markers. Molecular genetics and genomics: MGG **268**, 543–552.

Chen T, Li E. 2004. Structure and function of eukaryotic DNA methyltransferases. Current topics in developmental biology **60**, 55–89.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Molecular ecology **21**, 1548–1566.

Crutzen PJ. 2002. Geology of mankind. Nature **415**, 23–23.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401**, 157–161.

Danyluk J, Kane N, Breton G, Limin A, Fowler D, Sarhan F. 2003. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiology **132**, 1849–1860.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Dawson JC, Goldringer I. 2012. Breeding for genetically diverse populations: variety mixtures and evolutionary populations. Organic Crop Breeding, 77–98.

Day T, Bonduriansky R. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. The American naturalist **178**, E18–36.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature **418**, 700–707.

Elias M, McKey D, Panaud O, Anstett M, Robert T. 2001. Traditional management of cassava morphological and genetic diversity by the Makushi Amerindians (Guyana, South America): Perspectives for on-farm conservation of crop genetic resources. Euphytica **120**, 143–157.

Elias M, Rival L, Mckey D. 2000. Perception and management of cassava (Manihot esculenta Crantz) diversity among Makushi Amerindians of Guyana (South America). Journal of Ethnobiology **20**, 239–265.

Emperaire L, Peroni N. 2007. Traditional Management of Agrobiodiversity in Brazil: A Case Study of Manioc. Human Ecology **35**, 761–768.

Engelmann F, Engels JMM. 2002. Technologies and strategies for ex situ conservation. Managing plant genetic diversity, 89–103.

Engels J. 2002. *Managing plant genetic diversity [electronic resource]*. CABI.

Enjalbert J, Dawson JC, Paillard S, Rhoné B, Rousselle Y, Thomas M, Goldringer I. 2011. Dynamic management of crop diversity: From an experimental approach to on-farm conservation. Comptes Rendus Biologies **334**, 458–468.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences of the United States of America **95**, 4441–4446.

Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. Benjamin Cummings.

FAO. 2010. *The second report on the state of the world's plant genetic resources for food and agriculture.* Rome: FAO.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. Science (New York, N.Y.) **330**, 622–627.

Finckh MR. 2008. Integration of breeding and technology into diversification strategies for disease control in modern agriculture. European Journal of Plant Pathology **121**, 399–409.

Fisher RA. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh **52**, 399–433.

Fisher RA. 1930*a*. *The Genetical Theory Of Natural Selection*. At The Clarendon Press.

Fisher RA. 1930*b*. Genetics, Mathematics, and Natural Selection. Nature **126**, 805–806.

Flatscher R, Frajman B, Schönswetter P, Paun O. 2012. Environmental Heterogeneity and Phenotypic Divergence: Can Heritable Epigenetic Variation Aid Speciation? Genetics Research International **2012**.

Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, Holling CS. 2004. Regime shifts, resislience, and biodiversity in ecosystem management. Annual Review of Ecology, Evolution, and Systematics **35**, 557–581.

Frison EA, Cherfas J, Hodgkin T. 2011. Agricultural Biodiversity Is Essential for a Sustainable Improvement in Food and Nutrition Security. Sustainability **3**, 238–253.

Fu D, Szucs P, Yan L, Helguera M, Skinner JS, von Zitzewitz J, Hayes PM, Dubcovsky J. 2005. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. Molecular Genetics and Genomics: MGG **273**, 54–65.

Gao L, Geng Y, Li B, Chen J, Yang J. 2010. Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. Plant, cell & environment **33**, 1820–1827.

Gepts P, Bettinger R, Brush S, Damania A, Famula T, McGuire P, Qualset C. 2012. Introduction: The Domestication of Plants and Animals: Ten Unanswered Questions. Biodiversity in Agriculture. Cambridge University Press, .

Giménez-Benavides L, García-Camacho R, Iriondo JM, Escudero A. 2011. Selection on flowering time in Mediterranean high-mountain plants under global warming. Evolutionary Ecology **25**, 777–794.

Goldringer I. 2006. Rapid Differentiation of Experimental Populations of Wheat for Heading Time in Response to Local Climatic Conditions. Annals of Botany **98**, 805–817.

Gourdji SM, Mathews KL, Reynolds M, Crossa J, Lobell DB. 2013. An assessment of wheat yield sensitivity and breeding gains in hot environments. Proceedings. Biological sciences / The Royal Society **280**, 20122190.

Gunderson L. 2000. Ecological resilience - in theory and application. Annual Review of Eocology and Systematics **31**, 425–439.

Hall MC, Willis JH. 2006. Divergent Selection on Flowering Time Contributes to Local Adaptation in Mimulus Guttatus Populations. Evolution **60**, 2466–2477.

Hammer K, Teklu Y. 2008. Plant Genetic Resources: Selected Issues from Genetic Erosion to Genetic Engineering. Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS) **109**, 15–50.

Hanocq E, Laperche A, Jaminon O, Lainé A-L, Le Gouis J. 2007. Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **114**, 569–584.

Haudricourt AG. 1964. Nature et culture dans la civilisation de l'igname : l'origine des clones et des clans. L'Homme **4**, 93–104.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S, *et al.* 2007. Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication. Molecular Biology and Evolution **24**, 1506–1517.

Herrera CM, Bazaga P. 2010. Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. The New phytologist **187**, 867–876.

Herrera CM, Bazaga P. 2011. Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. Molecular ecology **20**, 1675–1688.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. Nature **470**, 479–485.

Holliday R. 1994. Epigenetics: An overview. Developmental Genetics **15**, 453–457.

Holt RD, Gomulkiewicz R. 1997. How does immigration influence local adaptation? A reexamination of a familiar paradigm. American Naturalist **149**, 563–572.

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, *et al.* 2005. EFFECTS OF BIODIVERSITY ON ECOSYSTEM FUNCTIONING: A CONSENSUS OF CURRENT KNOWLEDGE. Ecological Monographs **75**, 3–35.

Huang J, Pray C, Rozelle S. 2002. Enhancing the crops to feed the poor. Nature **418**, 678–684.

Jablonka E. 1995. *Epigenetic Inheritance and Evolution: Lamarckian Dimension*. Oxford University Press.

Jablonka E, Lamb MJ. 1989. The inheritance of acquired epigenetic variations. Journal of theoretical biology **139**, 69–83.

Jablonka E, Lamb MJ. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, And Symbolic Variation in the History of Life*. A Bradford Book.

Jablonka E, Raz G. 2009. TRANSGENERATIONAL EPIGENETIC INHERITANCE: PREVALENCE, MECHANISMS, AND IMPLICATIONS FOR THE STUDY OF HEREDITY AND EVOLUTION. QUARTERLY REVIEW OF BIOLOGY **84**, 131–176.

Jackson LE, Pascual U, Hodgkin T. 2007. Utilizing and conserving agrobiodiversity in agricultural landscapes. Agriculture, Ecosystems & Environment **121**, 196–210.

Jarvis DI, Padoch C, Cooper HD. 2007. Biodiversity, agriculture and ecosystem services - Google Scholar. Managing Biodiversity in Agricultural Ecosystems. Jarvis D. I., Padoch C and Copper H. D., 1–12.

Jenuwein T, Allis CD. 2001. Translating the histone code. Science (New York, N.Y.) **293**, 1074–1080.

Johannes F, Colot V, Jansen RC. 2008. Epigenome dynamics: a quantitative genetics perspective. Nature reviews. Genetics **9**, 883–890.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**, 94.

Kamran A, Iqbal M, Navabi A, Randhawa H, Pozniak C, Spaner D. 2013. Earliness per se QTLs and their interaction with the photoperiod insensitive allele Ppd-D1a in the Cutler × AC Barrie spring wheat population. Theoretical and Applied Genetics **126**, 1965–1976.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in Arabidopsis thaliana. Annual review of plant biology **55**, 141–172.

Kouzarides T. 2007. Chromatin modifications and their function. Cell **128**, 693–705.

Kremer A, Le Corre V. 2012. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. Heredity **108**, 375–385.

Kuchel H, Hollamby G, Langridge P, Williams K, Jefferies SP. 2006. Identification of genetic loci associated with ear-emergence in bread wheat. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **113**, 1103–1112.

Lachmann M, Jablonka E. 1996. The inheritance of phenotypes: an adaptation to fluctuating environments. Journal of theoretical biology **181**, 1–9.

Lammerts van Bueren E, Backes G, Vriend H, Østergård H. 2010*a*. The role of molecular markers and marker assisted selection in breeding for organic agriculture. Euphytica **175**, 51–64.

Lammerts van Bueren ET, Jones SS, Tamm L, Murphy KM, Myers JR, Leifert C, Messmer MM. 2010*b*. The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. NJAS - Wageningen Journal of Life Sciences **DOI:10.1016/j.njas.2010.04.001**.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics **11**, 204–220.

Lee J-S, Smith E, Shilatifard A. 2010. The Language of Histone Crosstalk. Cell **142**, 682–685.

Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, *et al.* 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC Genomics **13**, 300.

Loskutov IG. 2001. Influence of vernalization and photoperiod to the vegetation period of wild species of oats (Avena spp.). Euphytica **117**, 125–131.

Louette D, Smale M. 2000. Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica **113**, 25–41.

Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J. 2005. Regulation of VRN-1 Vernalization Genes in Normal and Transgenic Polyploid Wheat. Plant Physiol. **138**, 2364–2373.

Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature genetics **38**, 948–952.

Maxted N. 2012. *Agrobiodiversity Conservation*. CABI.

Maxted N, Ford-Lloyd B, Hawkes JG. 1997. Complementary conservation strategies. Plant genetic conservation, the in situ approach. Chapman and Hall, London, 15–39.

Maxted N, Guarino L, Myer L, Chiwona EA. 2002. Towards a methodology for on-farm conservation of plant genetic resources. Genetic Resources and Crop Evolution **49**, 31–46.

McKay JK, Latta RG. 2002. Adaptive population divergence: markers, QTL and traits. Trends in Ecology & Evolution **17**, 285–291.

Murphy K, Campbell K, Lyon S, Jones S. 2007. Evidence of varietal adaptation to organic farming systems. Field Crops Research **102**, 172–177.

Nei M. 1972. Genetic distances between populations. Am. Nat. **106**, 283–292.

Olesen JE, Trnka M, Kersebaum KC, Skjelvåg AO, Seguin B, Peltonen-Sainio P, Rossi F, Kozyra J, Micale F. 2011. Impacts and adaptation of European crop production systems to climate change. European Journal of Agronomy **34**, 96–112.

Pál C, Miklós I. 1999. Epigenetic Inheritance, Genetic Assimilation and Speciation. Journal of Theoretical Biology **200**, 19–37.

Parzies HK, Spoor W, Ennos RA. 2000. Genetic diversity of barley landrace accessions (Hordeum vulgare ssp. vulgare) conserved for different lengths of time in ex situ gene banks. Heredity **84 (Pt 4)**, 476–486.

Paun O, Bateman RM, Fay MF, Hedrén M, Civeyrel L, Chase MW. 2010. Stable Epigenetic Effects Impact Adaptation in Allopolyploid Orchids (Dactylorhiza: Orchidaceae). Molecular Biology and Evolution **27**, 2465 – 2473.

Paun O, Bateman RM, Fay MF, Luna JA, Moat J, Hedrén M, Chase MW. 2011. Altered gene expression and ecological divergence in sibling allopolyploids of Dactylorhiza (Orchidaceae). BMC Evolutionary Biology **11**, 113.

Phillips SL, Wolfe MS. 2005. Evolutionary plant breeding for low input systems. JOURNAL OF AGRICULTURAL SCIENCE-CAMBRIDGE- **143**, 245.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. Nature **457**, 843–848.

Raj S, Brautigam K, Hamanishi ET, Wilkins O, Thomas BR, Schroeder W, Mansfield SD, Plant AL, Campbell MM. 2011. Clone history shapes Populus drought responses. Proceedings of the National Academy of Sciences of the United States of America **108**, 12521–12526.

Rao NK. 2004. Plant genetic resources: advancing conservation and use through biotechnology. African Journal of biotechnology **3**, 136–145.

Rapp R, Wendel J. 2005. Epigenetics and plant evolution. NEW PHYTOLOGIST **168**, 81–91.

Raquin A-L, Brabant P, Rhoné B, Balfourier F, Leroy P, Goldringer I. 2008. Soft selective sweep near a gene that increases plant height in wheat. Molecular Ecology **17**, 741–756.

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. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **106**, 895–903.

Rhone B, Remoue 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.

Rhone B, Vitalis R, Goldringer I, Bonnin I. 2010. Evolution of Flowering Time in Experimental Wheat Populations: A Comprehensive Approach to Detect Genetic Signatures of Natural Selection. Evolution **64**, 2110–2125.

Richards EJ. 2009. Quantitative epigenetics: DNA sequence variation need not apply. Genes & development **23**, 1601–1605.

Richards C, Bossdorf O, Verhoeven K. 2010. Understanding natural epigenetic variation. NEW PHYTOLOGIST **187**, 562–564.

Rindos D. 1984. *The origins of agriculture: an evolutionary perspective*. Academic Press.

Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, *et al.* 2009. A safe operating space for humanity. Nature **461**, 472–475.

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, *et al.* 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO Journal **30**, 1928–1938.

Rozenfeld AF, Arnaud-Haond S, Hernández-García E, Eguíluz VM, Serrão EA, Duarte CM. 2008. Network analysis identifies weak and strong links in a metapopulation system. Proceedings of the National Academy of Sciences **105**, 18824–18829.

Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Molecular ecology **14**, 1163–1175.

Sandring S, Riihimäki M-A, 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.

Sasani S, Hemming M, Oliver S, Greenup A, Tavakkol-Afshari R, Mahfoozi S, Poustini K, Sharifi H, Dennis E, Peacock W, *et al.* 2009. The influence of vernalization and daylength on expression of flowering-time genes in the shoot apex and leaves of barley (Hordeum vulgare). Journal of Experimental Botany **60**, 2169–2178.

Smith M, Castillo F, Gomez F. 2001. Participatory plant breeding with maize in Mexico and Honduras. EUPHYTICA **122**, 551–565.

Soengas P, Cartea E, Lema M, Velasco P. 2008. Effect of regeneration procedures on the genetic integrity of Brassica oleracea accessions. Molecular Breeding **23**, 389–395.

Sokal RR, Michener CD. 1958. *A Statistical Method for Evaluating Systematic Relationships*. University of Kansas.

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 / National Research Council Canada = Génome / Conseil national de recherches Canada **43**, 487–494.

Takuno S, Gaut BS. 2013. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proceedings of the National Academy of Sciences of the United States of America **110**, 1797–1802.

Tal O, Kisdi E, Jablonka E. 2010. Epigenetic Contribution to Covariance Between Relatives. Genetics **184**, 1037–1050.

Thomas M. 2011. Gestion dynamique à la ferme de l'agrobiodiversité: relation entre la structure des populations de blé tendre et les pratiques humaines.

Thomas M, Dawson JC, Goldringer I, Bonneuil C. 2011. Seed exchanges, a key to analyze crop diversity dynamics in farmer-led on-farm conservation. Genetic Resources and Crop Evolution **58**, 321–338.

Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics **169**, 1589–1599.

Trevaskis B. 2010. The central role of the VERNALIZATION1 gene in the vernalization response of cereals. Functional Plant Biology **37**, 479–487.

Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES. 2003. MADS box genes control vernalization-induced flowering in cereals. Proceedings of the National Academy of Sciences of the United States of America **100**, 13099–13104.

Trevaskis B, Hemming M, Peacock W, Dennis E. 2006. HvVRN2 responds to daylength, whereas HvVRN1 is regulated by vernalization and developmental status. Plant Physiology **140**, 1397–1405.

Turelli M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theoretical population biology **25**, 138–193.

UNCED. 1992. *Convention on biological diversity*. Department of Public Information, United Nations, New York.

Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, *et al.* 2007. Epigenetic Natural Variation in Arabidopsis thaliana. PLoS Biology **5**, e174.

Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. The New phytologist **185**, 1108–1118.

Via S, Lande R. 1985. Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity. Evolution **39**, 505.

Vijg J, Suh Y. 2013. Genome Instability and Aging. Annual Review of Physiology **75**, 645–668.

Visser ME. 2008. Keeping up with a warming world; assessing the rate of adaptation to climate change. Proceedings of the Royal Society B: Biological Sciences **275**, 649–659.

Wolfe MS. 2000. Crop strength through diversity. Nature **406**, 681–682.

Wolfe MS, Baresel JP, Desclaux D, Goldringer I, Hoad S, Kovacs G, Löschenberger F, Miedaner T, Østergård H, Lammerts van Bueren ET. 2008. Developments in breeding cereals for organic agriculture. Euphytica **163**, 323–346.

Wood D, Lenné JM. 1997. The conservation of agrobiodiversity on-farm: questioning the emerging paradigm. Biodiversity and Conservation **6**, 109–129.

Wood D, Lenné JM. 1999. *Agrobiodiversity: characterization, utilization, and management*. CABI Pub.

Lev-Yadun S, Gopher A, Abbo S. 2000. The Cradle of Agriculture. Science **288**, 1602–1603.

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J. 2006. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proceedings of the National Academy of Sciences **103**, 19581–19586.

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. 2004. The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. Science **303**, 1640–1644.

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 of the United States of America **100**, 6263–6268.

Yi C, Zhang S, Liu X, Bui HT, Hong Y. 2010. Does epigenetic polymorphism contribute to phenotypic variances in Jatropha curcas L.? BMC Plant Biology **10**, 259.

Zhang Y-Y, Fischer M, Colot V, Bossdorf O. 2013. Epigenetic variation creates potential for evolution of plant phenotypic plasticity. The New phytologist **197**, 314–322.

Zhang M, Kimatu JN, Xu K, Liu B. 2010. DNA cytosine methylation in plant development. Journal of Genetics and Genomics = Yi Chuan Xue Bao **37**, 1–12.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, *et al.* 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell **126**, 1189–1201.

Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, Fan J, Yang S, Hu L, Leung H, *et al.* 2000. Genetic diversity and disease control in rice. Nature **406**, 718–722.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2006. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics **39**, 61–69.

Nei M, Tajima F. 1981. Genetic Drift and Estimation of Effective Population Size. Genetics **98**, 625–640.

Newton AC, Akar T, Baresel JP, Bebeli PJ, Bettencourt E, Bladenopoulos KV, Czembor JH, Fasoula DA, Katsiotis A, Koutis K, *et al.* 2010. Cereal landraces for sustainable agriculture. A review. Agronomy for Sustainable Development **30**, 237–269.

Newton AC, Begg GS, Swanston JS. 2009. Deployment of diversity for enhanced crop function. Annals of Applied Biology **154**, 309–322.

Olesen JE, Trnka M, Kersebaum KC, Skjelvåg AO, Seguin B, Peltonen-Sainio P, Rossi F, Kozyra J, Micale F. 2011. Impacts and adaptation of European crop production systems to climate change. European Journal of Agronomy **34**, 96–112.

Osman A, Chable V. 2009. Inventory of initiatives on seeds of landraces in Europe. Journal of Agriculture and Environment for International Development **103**, 95–130.

Ostergaard H, Finckh MR, Fontaine L, Goldringer I, Hoad SP, Kristensen K, van Bueren ET., Mascher F, Munk L, Wolfe MS. 2009. Time for a shift in crop production: embracing complexity through diversity at all levels. Journal of the Science of Food and Agriculture **89**, 1439–1445.

Parzies HK, Spoor W, Ennos RA. 2000. Genetic diversity of barley landrace accessions (Hordeum vulgare ssp. vulgare) conserved for different lengths of time in ex situ gene banks. Heredity **84 (Pt 4)**, 476–486.

Perrier X, Jacquemoud-Collet, JP 2006. **DARwin** software http://darwin.cirad.fr/

Pistorius R. 1997. *Scientists, Plants and Politics: A History of the Plant Genetic Resources Movement*. Bioversity International.

Plucknett DL. 1987. *Gene banks and the world's food*. Princeton University Press.

R Development Core Team. 2009. *{R: A language and environment for statistical computing}*. Vienna, Austria: R Foundation for Statistical Computing.

Rhone B, Vitalis R, Goldringer I, Bonnin I. 2010. Evolution of Flowering Time in Experimental Wheat Populations: A Comprehensive Approach to Detect Genetic Signatures of Natural Selection. Evolution **64**, 2110–2125.

Rice EB, Smith ME, Mitchell SE, Kresovich S. 2006. Conservation and Change: A Comparison of In situ and Ex situ Conservation of Jala Maize Germplasm. Crop Science **46**, 428.

Roussel V, Koenig J, Beckert M, Balfourier F. 2004. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. TAG Theoretical and Applied Genetics **108**, 920–930.

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F. 2005. SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **111**, 162–170.

Rousset M, Bonnin I, Remoué C, Falque M, Rhoné B, Veyrieras J-B, Madur D, Murigneux A, Balfourier F, Gouis J, *et al.* 2011. Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (Triticum aestivum L.). Theoretical and Applied Genetics **123**, 907–926.

Russell JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young G, Powell W. 2000. A retrospective analysis of spring barley germplasm development from `foundation genotypes' to currently successful cultivars. Molecular Breeding **6**, 553–568.

Serpolay E, Dawson JC, Chable V, Bueren EL, Osman A, Pino S, Silveri D, Goldringer I. 2011. Diversity of different farmer and modern wheat varieties cultivated in contrasting organic farming conditions in western Europe and implications for European seed and variety legislation. Organic Agriculture **1**, 127–145.

Simmonds NW. 1962. Variability in Crop Plants, Its Use and Conservation. Biological Reviews **37**, 422–465.

Soleri D, Smith SE. 1995. Morphological and phenological comparisons of two Hopi maize varieties conserved in situ and ex situ. Economic Botany **49**, 56–77.

Stephens M, Smith NJ, Donnelly P. 2001. A New Statistical Method for Haplotype Reconstruction from Population Data. The American Journal of Human Genetics **68**, 978–989.

Su Z, Hao C, Wang L, Dong Y, Zhang X. 2011. Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **122**, 211–223.

Sun J-C, Cao G-L, Ma J, Chen Y-F, Han L-Z. 2012. Comparative genetic structure within single-origin pairs of rice (Oryza sativa L.) landraces from in situ and ex situ conservation programs in Yunnan of China using microsatellite markers. Genetic Resources and Crop Evolution **59**, 1611–1623.

Thomas M. 2011. Gestion dynamique à la ferme de l'agrobiodiversité: relation entre la structure des populations de blé tendre et les pratiques humaines.

Thomas M, Dawson JC, Goldringer I, Bonneuil C. 2011. Seed exchanges, a key to analyze crop diversity dynamics in farmer-led on-farm conservation. Genetic Resources and Crop Evolution **58**, 321–338.

Thomas M, Demeulenaere E, Dawson JC, Khan AR, Galic N, Jouanne-Pin S, Remoue C, Bonneuil C, Goldringer I. 2012. On-farm dynamic management of genetic diversity: the impact of seed diffusions and seed saving practices on a population-variety of bread wheat. Evolutionary Applications **5**, 779–795.

Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics **169**, 1589–1599.

Tin HQ, Berg T, Bjørnstad Å. 2001. Diversity and adaptation in rice varieties under static (ex situ) and dynamic (in situ) management. Euphytica **122**, 491–502.

Torricelli R, Tiranti B, Spataro G, Castellini G, Albertini E, Falcinelli M, Negri V. 2013. Differentiation and structure of an Italian landrace of celery (Apium graveolens L.): inferences for on farm conservation. Genetic Resources and Crop Evolution **60**, 995–1006.

UNCED. 1992. Convention on Biological Diversity. In Proceedings of the *United Nation Conference on Environment and Development, Rio de Janeiro, Brazil*. Department of Public Information, United Nations, New York.

Waples RS. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics **121**, 379–391.

White JW, Herndl M, Hunt LA, Payne TS, Hoogenboom G. 2008. Simulation-Based Analysis of Effects of Vrn and Ppd Loci on Flowering in Wheat. Crop Science **48**, 678–687.

Wilhelm EP, Turner AS, Laurie DA. 2008. Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (Triticum durum Desf.). Theoretical and Applied Genetics **118**, 285–294.

Wolfe MS, Baresel JP, Desclaux D, Goldringer I, Hoad S, Kovacs G, Löschenberger F, Miedaner T, Østergård H, Bueren ETL van. 2008. Developments in breeding cereals for organic agriculture. Euphytica **163**, 323–346.

Supplimentory Table 2.1: Candidate gene markers with their codes and with their expected gene

Lab marker name	Marker	Pos	Chr	Gene name	Analysis Code
CA_2045_41	wsnp_Ex_c1563_2987002	156	4A	PHYA	CA1
CA_2307_43	wsnp_Ex_c18382_27210656	82	6B	ZTL	CA2
CA_3677_25	wsnp_Ex_c39304_46635517	84	6B	VIL2	CA3
CA_4049_34	wsnp_Ex_c4921_8764088	186	5A	CO3	CA4
CA_44_33	wsnp_BE403956B_Ta_2_3	59	1B	SMZ	CA5
CA_4509_32	wsnp_Ex_c645_1273901	102	1A	Vrn1B	CA6
CA_4699_22	wsnp_Ex_c7546_12900094	256	6A	Vrn-1B	CA7
CA_4805_29	wsnp_Ex_c8424_14192191	89	5A	FT	CA8
CA_4872_52	wsnp_Ex_c9063_15093396	114	1B	SMZ	CA9
CA_4916_50	wsnp_Ex_c9440_15657149	49	4B	CO1	CA10
CA_5042_26	wsnp_Ex_rep_c102044_87296690	84	6B	VIL2	CA11
CA_5269_42	wsnp_Ex_rep_c66600_64897324	156	4A	PHYA	CA12
CA_5396_37	wsnp_Ex_rep_c67404_65986980	115	3B	TaGI3	CA13
CA_5440_36	wsnp_Ex_rep_c67690_66354931	224	5B	CO4	CA14
CA_5656_38	wsnp_Ex_rep_c69901_68864080	106	6A	Vrn1B	CA15
CA_5860_30	wsnp_JD_c15333_14824351	155	7A	CO1	CA16
CA_6412_24	wsnp_Ku_c1102_2211433	232	5A	LDD-A	CA17
CA_6574_40	wsnp_Ku_c15816_24541712	175	5A	TaHd1A	CA18
CA_7108_23	wsnp_Ku_c48167_54427241	201	3B	CO1	CA19
CA_750_45	wsnp_CAP11_c3346_1639010	182	4A	SMZ	CA20
CA_7643_35	wsnp_Ra_c16053_24607526	203	3A	SOC1	CA21
CA_7895_46	wsnp_Ra_c3766_6947230	82	6B	ZTL	CA22
CA_923_27	wsnp_CAP12_c1461_744121	142	7A	CO4	CA23
CA_CH_47				ZTL	CA24
COAB_1			5B	COAB	CA25
FTA_2			7A	FT-A	CA26
PPD-A1-CADE_14			2A	PPD-A1	CA27
PPD-B1-SNP_CT_18			2B	PPD-B1	CA28
PPD-B1-SNP_GC_19			2B	PPD-B1	CA29
PPD-D1PROM_6			2D	PPD-D1	CA30
TAGW2_15			6A	TAGW2	CA31
VRN1A-EX8_20			5A	VRN-1A	CA32
VRN-1DIN1_7				VRN-D1	CA33
CA_6905_28					CA34

Supplementary table 2.1 continued

Lab Marker name	Marker	Pos	Chr	Gene name	Code
CA_2164_53	wsnp_Ex_c16720_25268525	58	1D	Vrn1D	CA35
CA_3692_54	wsnp_Ex_c4612_8254533	178	1A		CA36
PPD-A1-GS100_12			2A	PPD-A1	CA37
PPD-A1-GS105_13			2A	PPD-A1	CA38
PPD-D1EX8_5			2D	PPD-D1	CA39
RHT-D1_17			4D	RHT-D1	CA40
VRN1APR4/5_10			5A	VRN1A	CA41
VRN1BIN1_11			5B	VRN1B	CA42
CA_4974_49	wsnp_Ex_c9872_16271161	95	1A	CDF1	CA43
RHT-B1_16			4B	RHT-B1	CA44
CA_7180_21	wsnp_Ku_c5623_9966516	151	6B	LDD-A	CA45
FTD_3			7D	FTD	CA46
LDDB_4			3B	LDD-B	CA47
CA_7896_31	wsnp_Ra_c3766_6947263	82	6B	ZTL	CA48

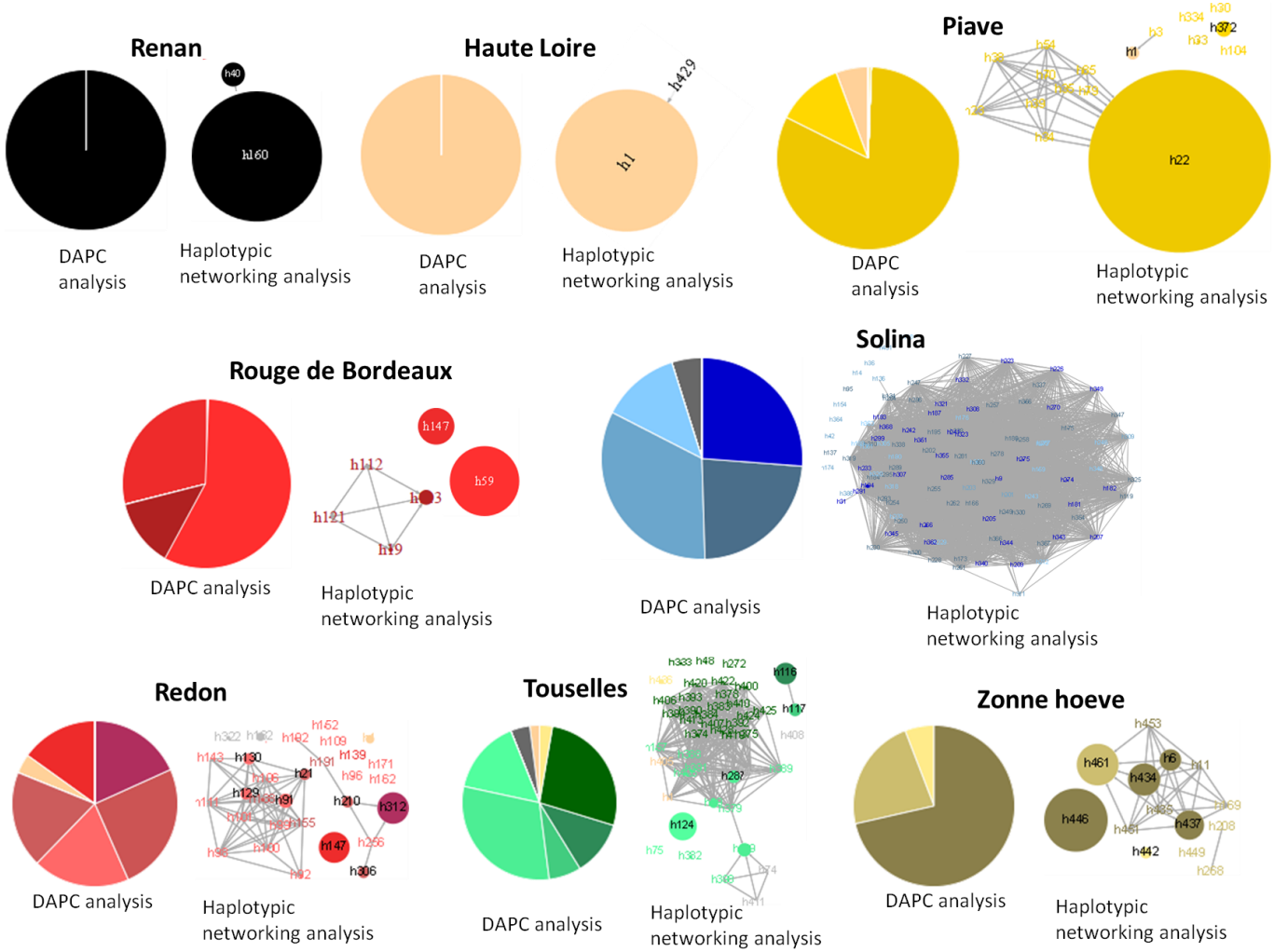
Where blue color shows monomorphic markers whereas red color shows bad quality markers

Supplementary Table 2.2: neutral markers with their codes

lab marker name	SNP Index	Marker	Pos	Chr	Code
NE_4816_27	4816	wsnp_Ex_c8588_14419007	123.8	1A	NE13
NE_578_5	578	wsnp_BG606986A_Ta_2_4	171.25	1A	NE21
NE_605_7	605	wsnp_BM140362A_Ta_2_2	143.88	1A	NE23
NE_5083_30	5083	wsnp_Ex_rep_c103087_88123733	111.3	1A	NE29
NE_1644_49	1644	wsnp_Ex_c1255_2411550	273.91	1A	NE33
NE_513_4	513	wsnp_BF484606A_Ta_2_3	54	1A	NE40
NE_5228_60	5228	wsnp_Ex_rep_c66389_64588992	117.75	1B	NE35
NE_7533_44	7533	wsnp_Ra_c1020_2062200	27.16	1D	NE3
NE_7547_45	7547	wsnp_Ra_c107797_91270622	74.39	2A	NE32
NE_143_1	143	wsnp_BE443995B_Ta_2_2	115.15	3A	NE38
NE_7519_89	7519	wsnp_Ku_rep_c73198_72796386	148.45	3B	NE4
NE_210_3	210	wsnp_BE489326B_Ta_2_1	124.4	3B	NE24
NE_6919_40	6919	wsnp_Ku_c33335_42844594	102.73	3B	NE31
NE_6485_37	6485	wsnp_Ku_c13204_21105694	0.55	3D	NE9
NE_4662_51	4662	wsnp_Ex_c7362_12622736	32.28	4B	NE14
NE_4465_58	4465	wsnp_Ex_c62701_62229607	245.31	5A	NE15
NE_7135_62	7135	wsnp_Ku_c51039_56457361	170.88	5A	NE36
NE_7507_69	7507	wsnp_Ku_rep_c72211_71920520	181.63	5B	NE5
NE_7471_66	7471	wsnp_Ku_rep_c70220_69775367	153.72	5B	NE6
NE_6366_74	6366	wsnp_JG_c625_379570	153.72	5B	NE10
NE_3630_24	3630	wsnp_Ex_c38105_45710671	224.43	5B	NE16
NE_3044_79	3044	wsnp_Ex_c26312_35558700	153.72	5B	NE17
NE_1460_15	1460	wsnp_Ex_c11265_18216936	90.92	5B	NE19
NE_987_48	987	wsnp_CAP12_c7952_3403722	171.58	5B	NE22
NE_6902_38	6902	wsnp_Ku_c3151_5892200	30.13	5B	NE30
NE_7177_42	7177	wsnp_Ku_c55961_59662821	6.87	5D1	NE7
NE_2821_20	2821	wsnp_Ex_c23618_32855041	6.87	5D1	NE26
NE_2366_19	2366	wsnp_Ex_c18965_27868480	106.39	6A	NE18
NE_4929_28	4929	wsnp_Ex_c9502_15748469	137.04	6A	NE27
NE_4961_29	4961	wsnp_Ex_c9763_16125630	43.61	6A	NE28
NE_7930_67	7930	wsnp_Ra_c4254_7755493	81.5	6B	NE2
NE_5666_52	5666	wsnp_Ex_rep_c70036_68988728	218.62	6B	NE12
NE_597_6	597	wsnp_BM136727B_Ta_2_6	151.4	6B	NE41
NE_618_8	618	wsnp_BQ161779B_Ta_2_4	151.53	6B	NE42

lab marker name	SNP Index	Marker	Pos	Chr	Code
NE_7005_41	7005	wsnp_Ku_c3929_7189422	10.08	7A	NE8
NE_5904_36	5904	wsnp_JD_c19925_17854742	10.08	7A	NE11
NE_5912_59	5912	wsnp_JD_c20555_18262260	98.15	7A	NE34
NE_832_70	832	wsnp_CAP11_rep_c4027_1902057	106.13	7B	NE1
NE_2353_18	2353	wsnp_Ex_c18800_27681277	78.86	7B	NE20
NE_5071_86	5071	wsnp_Ex_rep_c102707_87814407	136.4	7B	NE37
NE_182_2	182	wsnp_BE445506B_Ta_2_4	260.38	7B	NE39
NE_2328_17	2328	wsnp_Ex_c18616_27481826			NE25
NE_2180_16	2180	wsnp_Ex_c16963_25554152	181.63	5B	
NE_4113_26	4113	wsnp_Ex_c5185_9189184	29.71	5D2	
NE_8377_75	8377	wsnp_RFL_Contig2729_2446041	197.66	2A	

Where red color shows bad quality markers



Supplementary Figure 2.1: Comparison between the Genetic groups obtained by DAPC and haplotypic network analysis

Chapter 3

Vernalization treatment induces site-specific DNA hypermethylation at the VERNALIZATION-A1 (*VRN-A1*) locus in hexaploid winter wheat

At the interface between genotype and environment, the overall rate of epimutations is often much higher than that of genetic mutations (Tal *et al.*, 2010), resulting in a more dynamic level of variation (Flatscher *et al.*, 2012). For my PhD, we were interested in evaluating the response of populations cultivated in contrasting environments in a short period of time (three years). Because chromatin conformation allows a readout of the genome information (and therefore acts one step further towards phenotype) and is more dynamic than DNA sequence information, we were interested in evaluating the amount of epigenetic variation present in our populations, to decipher whether it could act as major driving force in the rapid adaptive processes observed. Among chromatin marks (such as DNA methylation, histone modifications and histone variants), we chose to focus on DNA methylation, which is stable and offers easy development of markers for population studies. The first step towards developping epigenetic markers for population studies was to choose a gene which causes a phenotype and is regulated through changes in chromatin structure. Moreover, because the populations used in this study belonged to regions of origin with contrasting environmental conditions and are expected to behave differently in these divergent environments, we were seeking for a gene that would also respond to environmental changes.

Flowering time is a strongly heritable trait that is of main importance for plants. Flowering time can be selected for in different environments and therefore might lead to the adaptation of populations to local environments (Loskutov, 2001; Hall and Willis, 2006; Sandring et al., 2007; Giménez-Benavides et al., 2011). Flowering time involves three different but interconnected pathways (*i.e.*, vernalization, photoperiod and earliness *per se*).

Among these, vernalization involves mainly three genes: *VERNALIZATION1* (*VRN1), VERNALIZATION2* (*VRN2)* and FLOWERING TIME (*FT). VRN1* is a floral activator gene which is central in the vernalization pathway (Trevaskis, 2010). It down regulates the floral repressor *VRN2* and interacts with the floral activator *FT* to accelerate subsequent floral development. In winter cereals, expression of *VRN1* is induced by cold treatment, is maintained when this cold treatment is released, and is reset in the next generation (Trevaskis *et al.*, 2003, 2006; Yan *et al.*, 2003; Danyluk *et al.*, 2003; Loukoianov *et al.*, 2005; Sasani *et al.*, 2009), characteristics that indicate the possibility of an epigenetic regulation. For these reasons, *VRN1* qualified as a good candidate gene for use in populational studies. As common wheat is

hexaploid, it contains three copies of this gene (referred to as *VRN-A1*, *VRN-B1* and *VRN-D1*, A, B and D representing the three genome of wheat). Among those, *VRN-A1* has the strongest effect compared to *VRN-B1* and *VRN-D1* and was therefore selected for this study (Trevaskis *et al.*, 2003; Loukoianov *et al.*, 2005).

Due to the absence of any prior information about behaviour of DNA methylation at this gene, we first prerequisite for marker development was to study the DNA methylation pattern in the *VRN-A1* gene and investigate its response to vernalization treatment. This is what I investigated during the first two years of my PhD and the following chapter presents the findings of this study in the form of a research article.

Vernalization treatment induces site-specific DNA hypermethylation at the *VERNALIZATION-A1 (VRN-A1)* locus in hexaploid winter wheat

Abdul Rehman Khan[1], Jérôme Enjalbert[1], Anne-Charlotte Marsollier[1], Agnès Rousselet[1], Isabelle Goldringer[1], Clémentine Vitte[2*]

[1] INRA, UMR de Génétique Végétale, F-91190 Gif sur Yvette, France

[2] CNRS, UMR de Génétique Végétale, F-91190 Gif sur Yvette, France

*Corresponding author. Telephone: +33 1 69 33 23 57; Fax : +33 1 69 33 23 40

Email adresses :

khan@moulon.inra.fr

enjalbert@moulon.inra.fr

amarsoll@ens-cachan.fr

roussele@moulon.inra.fr

isa@moulon.inra.fr

vitte@moulon.inra.fr

Abstract

Background

Certain temperate species require prolonged exposure to low temperature to initiate transition from vegetative growth to flowering, a process known as vernalization. In wheat, winter cultivars require vernalization to initiate flowering, making vernalization requirement a trait of key importance in wheat agronomy. The genetic bases of vernalization response have been largely studied in wheat, leading to the characterization of a regulation pathway that involves the key gene *VERNALIZATION1 (VRN1)*. While previous studies in wheat and barley have revealed the functional role of histone modification in setting *VRN1* expression, other mechanisms might also be involved. Here, we were interested in determining whether the cold-induced expression of the wheat *VRN-A1* gene is associated with a change in DNA methylation.

Results

We provide the first DNA methylation analysis of the *VRN-A1* gene, and describe the existence of methylation at CG but also at non CG sites. While CG sites show a bell-shape profile typical of gene-body-methylation, non CG methylation is restricted to the large (8.5 kb) intron 1, in a region harboring fragments of transposable elements (TEs). Interestingly, cold induces a site-specific hypermethylation at these non CG sites. This increase in DNA methylation is transmitted through mitosis, and is reset to its original level after sexual reproduction.

Conclusions

These results demonstrate that *VRN-A1* has a particular DNA methylation pattern, exhibiting rapid shift within the life cycle of a winter wheat plant following exposure to particular environmental conditions. The finding that this shift occurs at non CG sites in a TE-rich region opens interesting questions onto the possible consequences of this type of methylation in gene expression.

Keywords

DNA methylation, non CG methylation, winter wheat, transposable element, *Triticum aestivum*, vernalization, cold, *VRN1*, intron, gene expression.

The rest of the article is not presented due to confidentiality reasons

Chapter 4

General Discussion

Approaches to study the short-term response to contrasting agro-climatic conditions

The wheat varieties used in this study comprise a set of landraces, mixtures, historic and modern varieties, each with a different conservation / management history. Since no prior information on the genetic structure of most of these varieties was available, we first characterized the fine genetic structure of these varieties, and then studied the *spatio-temporal* differentiation among different versions of each variety, to decipher the genetic and phenotypic response of these varieties in new and contrasting environments.

The genetic structure of these varieties was investigated using two complementary methods. First, we used a multivariate method (Discriminant Analysis of Principal Components) recently developed by (Jombart *et al.*, 2010). This method clusters individuals pooled from all samples of all varieties, thus allowing the identification of the different genetic groups. Some of these genetic groups were variety specific while others were shared between several varieties. This gave an overview of the within-variety diversity structure. Second, a method inspired from the network theory (Rozenfeld *et al.*, 2008), that has been developed in our team (Thomas, 2011) was applied on the dataset (see Chapter 2). Using the haploid phased multi-locus genotypes, so called haplotypic networks were developed by using this method on the complete data set, which allowed us to identify the major haplotypes, the finer structure of the genetic groups (as detected by DAPC) and the relatedness among these groups. Combination of these two methods allowed us to draw a clear picture of the fine genetic structure of the different wheat varieties, the level of within-variety genetic diversity, and how they are related to each other.

The *spatio-temporal* differentiation among the 2009 versions of all varieties was studied with conventional methods. First, we used the canonical analysis AFC (Factorial Analysis of Correspondences) to visualize the genetic differentiation among different populations in a three dimensional graph. Then, Analysis of MOlecular VAriance (AMOVA) allowed to quantify the overall genetic variation at different levels: among varieties, within-varieties among population, within-population. Then, genetic distances (Nei, 1972) between pairs of populations were estimated and a clustering was performed on the distance matrix to draw a classification tree using the UPGMA method (Sokal and Michener, 1958). The robustness of the classification obtained was then tested using bootstrap algorithms for genetic markers. Since the varieties used

in this study consisted of a set of landraces, mixtures, historic and modern varieties, each with a different conservation history, the level of structural variation and of within-variety diversity highly differed among varieties. Consideration of these factors was essential to properly understand the forces at work in shaping up the genetic differentiation pattern of these varieties due to contrasting agro-climatic conditions. Therefore, the observed pattern of *spatio-temporal* differentiation among different versions of each variety was supplemented by the insight about the fine genetic structure provided by the DAPC and haplotypic networking approaches to understand the genetic and phenotypic response of these varieties in new and contrasting environments.

Within-population genetic diversity in contrasting agro-climatic conditions

Biodiversity provides the raw material for evolution and adaptation of populations and species. In agricultural biodiversity, the within-population genetic diversity is of major importance. On one hand it can provide a buffering effect against the year-to-year variation of climate or biotic pressures and on the other hand diversity serves as a resource for the population to respond to selective pressures due to specific local conditions, thus allowing for local adaptation, particularly in the case where a population is introduced into a new location. Theoretically, the rate of adaptation is predominantly driven by the amount of available additive genetic variation at relevant adaptive traits and by the strength of environmental selection (Fisher, 1930*b*; Turelli, 1984). However, in natural conditions the interaction between demographic processes and evolutionary dynamics is very complex and understanding the genetic (or non-genetic) bases of the response to selection is not straightforward. While in Fisher theory of natural selection, it is assumed that fitness traits are highly complex, determined by a high number of loci showing alleles with small effects, the so-called Infinitesimal Model (Fisher, 1918), the advances in molecular biology, gene cloning and the numerous studies detecting QTLs (Quantitative Traits Loci) have shown that the variation of quantitative traits (and in particular of fitness traits) also relies on the variation at loci with larger effects or major genes in addition to multiple small gene effects (Bost *et al.*, 2001). Interactions among loci (epistasis) are also supposed to control part of the phenotype for complex traits and thus part of the adaptive response. Therefore, when considering the differentiation among sub-populations distributed in contrasted environments

and submitted to divergent selective pressures, there can be a large discrepancy between phenotypic differentiation at adaptive traits and differentiation at the QTLs underlying the traits (see Le Corre and Kremer, 2012 for a review). This is because part of the phenotypic differentiation is due to covariances among alleles at the underlying QTLs (McKay and Latta, 2002).

In this study, we had little information on the genetic variability of fitness traits involved in local adaptation, that was initially available for selection in the populations studied, but rather we had a good description of diversity at molecular markers (*i. e.*, Single Nucleotide Polymorphisms, SNPs) located in candidate genes or randomly in the genome. The impact of selection on neutral and gene diversity strongly depends on the structure of genetic diversity within and among the populations submitted to selection since it determines the extent of linkage disequilibrium among markers and genes controlling the adaptive traits. Therefore, as the set of varieties used in this study represent varying levels of within-population genetic diversity and structure (initial 2006 population), it allowed us to study the effect of different levels of within-population diversity on the local adaptation in the new environment in a short period of time (3 years) and in turn, to assess the impact of selection and genetic drift on the diversity.

Surprisingly, all varieties showed a significant differentiation among populations for the two phenotypic traits studied (plant height and heading date), although the range of variation of population means differed slightly from one variety to the other with a lower differentiation among populations for the modern varieties Renan and the landraces Haute-Loire and Solina.

Moreover, we found that the varieties with higher within-population diversity showed greater level of genetic differentiation at the neutral markers level after three generations of reproduction in contrasted environments than the varieties with low within-population diversity such as Renan and the landrace Haute-Loire. The highest differentiation was found for Redon, Rouge de Bordeaux, Piave and Touselle, followed by Zonne Hoeve. A similar pattern was obtained for differentiation at candidate gene markers. Thus, although phenotypic differentiation was not strongly correlated with the initial genetic diversity nor with the genetic differentiation, the populations that were phenotypically the most responsive were also the most diverse. The group structure of the varieties seemed to influence, to some extent, the genetic and phenotypic differentiation. For instance, Solina, one of the more genetically diverse varieties but composed

of less distinct groups with highly connected haplotypes within and among groups, showed moderate genetic and phenotypic differentiation. This suggests that Solina genetic diversity, which is derived from a long history of on-farm propagation, consists in a more continuous variation in genotypes and at quantitative traits, thus allowing for a slow response to environmental selective pressures together with the maintenance of within-population diversity and complexity. On the other hand, varieties composed of very distinct genetic groups such as Rouge de Bordeaux or Piave showed more drastic response that might be due to the selection of one group or the other. Finally, looking for temporal differentiation at individual candidate genes indicated that variation at these genes might, in some cases, explain the observed phenotypic response since several genes were detected as submitted to significantly large temporal change in frequency among which some of them were associated to phenotypic variation.

Interestingly, we have observed the differentiation in a short period of time (three years), while several studies of wheat dynamic management populations showed differentiation requiring a longer time such as (Rhone *et al.*, 2008) and Rhone *et al.*, (2010) where a significant spatial differentiation after the seventh generation was found, and Goldringer, (2006) where significant spatial differentiation was observed after ten years. This may be due to the more contrasted agro-climatic conditions in our study, to the difference in the genetic material of the dynamic management populations which was derived from multiple crosses among 16 parents and/or to larger genetic drift effects due to the smaller size of the plots in our study. As the methods used in these wheat dynamic management population studies did not target much on the genetic structure of these populations, we do not know if the haplotypes present within each population were highly connected like in Solina. Yet, we do expect that the high number of different parents and the four successive generations of crosses have led to limited linkage disequilibrium in the population and multiple connected haplotypes. Indeed, (Raquin *et al.*, 2008) found that the extend of linkage disequilibrium around a gene submitted to selection (the dwarfing gene *Rht1*) was quite low. This would then be in accordance with our observation of slow differentiation in Solina and support our hypothesis that presence of multiple connected haplotypes within the population, can lead to a buffering effect (i.e. to balance the effect of environmental variation) associated to lower differentiation.

Although the modern variety Renan showed a very low (even negligible) level of within-population diversity, indicating a highly fixed variety with almost no genetic differentiation among populations in 2009, significant phenotypic differentiation was found as well as differentiation at two specific candidate genes. The potential for phenotypic evolution of this variety might be due to some genetic residual variation that is kept unfixed in the commercially distributed seeds or to transmissible plasticity potentially controlled by methylation marks. This higher genetic or plastic adaptability of Renan might be the reason for its continuous use by many farmers in organic agriculture.

This study argues for the better adaptive potential of traditional varieties compared to modern varieties or to landraces conserved *ex situ,* especially in organic conditions where heterogeneous biotic and abiotic conditions cannot be balanced by high inputs. This emphasizes the need for either using heterogenous historic varieties or breeding for organic agriculture by keeping in mind the specific needs of organic agriculture systems.

These findings were consistent with the expected role of within-population diversity as an important prerequisite for local adaptation. Yet, part of the phenotypic response could not be explained by the genetic diversity we looked at, and thus there is a need for further analysis of diversity at more candidate genes, as well as epigenetic variation within and among varieties.

Towards epigenetic marker development to unravel the epigenetic variability

For my PhD, I was interested in evaluating the response of populations cultivated in contrasting environmental conditions in a short period of time (3 years). The rate of epimutations being often higher than that of genetic mutation (Tal *et al.*, 2010; Vijg and Suh, 2013), we wanted to investigate the impact of contrasting environments on the epigenetic variation of these populations. The first part of my PhD has therefore been dedicated to characterize the DNA methylation pattern of the *VRN-A1* locus as well as its response to vernalization treatment, as a first step to develop epigenetic markers on this gene.

As an outcome of this study, a region located in the intron 1 of the *VRN-A1* gene was shown to be hypermethylated in response to cold, at specific non-CG sites. This increase in DNA methylation is positively correlated with gene expression, suggesting a possible role of this

methylation in the regulation of the *VRN-A1* gene. Interestingly, deletions in intron 1 of *VRN-1* correlate with spring habits (Fu *et al.*, 2005), which supports the idea that intron 1 is involved in the gene regulation. The observation that this hypermethylation at specifically non-CG sites lies in a region that contains fragments of transposable elements suggests that the process involved in the silencing of these fragments may be involved. Interestingly, cold induces an increase in DNA methylation at these fragments, while a light global demethylation was observed at the genome wide level, thus suggesting that the TE fragments of this region may be regulated differently than the other TEs of the genome.

With our experiment only, it is difficult to establish whether this site-specific hypermethylation is involved in the regulation of the gene, or if it is a by-product of gene expression. Thus, additional information is needed to be able to use this region to develop markers for population studies.

In this study, we limited our experiment to two winter wheat genotypes that were submitted to one (mild) cold treatment. Therefore, to get better insights on the implication of the hypermethylation observed in the vernalization response, it would be interesting to analyse the level of methylation observed following a broad range of mild to more severe (*i.e.* longer time and/or lower temperature) cold treatments.

In parallel, testing the DNA methylation variation pattern in a larger number of genotypes with known and contrasted vernalization requirements would (i) increase the possibility of identifying other epialleles and (ii) allow to study the association between the DNA methylation level and the phenotypic response to cold in terms of earliness/flowering time. Altogether, this would help understanding the role of this site-specific hypermethylation in the wheat response to cold. If a good positive correlation is found, population analyses using DNA methylation markers developed for this region would be possible, and would allow to test whether DNA methylation changes have occurred in our wheat populations grown in contrasted environment, thus giving first clues on whether DNA methylation plays a role in wheat adaptation.

General Conclusion

This thesis has explored the short term response of European wheat populations (farmers' varieties) in contrasting agro-climatic conditions and how the genetic structure of these population influences this response. This study shows that genetic structure, especially the within-population diversity, greatly influences the response of populations when they are introduced to a relatively contrasting new environment. Interestingly, the conservation/management history plays an important role in shaping up the genetic structure/architecture of these populations and by extension the population response to environmental variability. The within-population diversity is observed to be lower in populations with *ex situ* conservation history whereas *in situ* conservation, *i.e.* on-farm conservation, tends to maintain and create within-variety genetic diversity in a dynamic and continuously evolving manner. Since both conservation approaches are complementary to each other, we propose the use of mixtures of related landraces to increase the genetic diversity while reintroducing the *ex situ* conserved varieties on farm, to take advantage of both conservation strategies. As for the short term response to contrasting environmental conditions, a certain level of genetic and phenotypic *spatio-temporal* differentiation is observed and is highly associated with the within-variety genetic diversity and structure. Populations with higher within-population diversity show greater genetic differentiation than those with lesser within-population diversity. Although phenotypic differentiation was not strongly correlated with the initial genetic diversity or with the genetic differentiation, the populations that were phenotypically the most responsive were also the most diverse. A significant phenotypic differentiation for varieties with very low genetic diversity has also been observed in this study, which gives indication of a possible role of epigenetic variation in the process of evolution.

Intrigued by the possibility of potential role of epigenetic variation in the adaptation of wheat varieties to varying environmental conditions, and with the objective of acquiring the ability to study this adaptive potential, we also analysed the epigenetic pattern (DNA methylation variation) of the *VRN-A1* gene in response to cold treatment, as a prerequisite information for the development of epigenetic markers. In addition to detecting gene body methylation across the *VRN-A1* gene, we identified a region within intron 1 that shows significant increase in DNA methylation in response to vernalization treatment that is positively correlated with the gene

expression. Although the role of this shift in gene regulation is still unclear due to time limitations in the thesis and the small number of genotypes analysed, this study will provide a good material towards future identification of new epialleles and the development of epigenetic markers to study the epigenetic variability of these populations. If natural epialleles can be identified, this work will pave the way into taking DNA methylation-based mechanisms into account in the process of breeding for local adaptation.

This study at large provides useful knowledge on the understanding of farmers' varieties evolutionary response to be used in the development of different breeding approaches for organic agriculture, taking into consideration of the importance of within-population diversity, to satisfactorily address the problems of organic agriculture.

References

Almekinders CJM, Elings A. 2001. Collaboration of farmers and breeders: Participatory crop improvement in perspective. Euphytica **122**, 425–438.

Altieri M, Anderson MK, Merrick L. 1987. Peasant Agriculture and the Conservation of Crop and Wild Plant Resources. Conservation Biology **1**, 49–58.

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.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **115**, 721–733.

Bellon MR. 1997. Understanding farmer seed and variety systems On-farm conservation as a process: an analysis of its components. USING DIVERSITY Enhancing and Maintaining Genetic Resources On-farm.

Bent AF. 2003. Crop diseases and stratigies for their control. Plants, Genes, and Crop Biotechnology. Jones & Bartlett Learning, .

Berger SL. 2007. The complex language of chromatin regulation during transcription. Nature **447**, 407–412.

Berthaud J, Clément JC, Emperaire L, Louette D, Pinton F, Sanou J, Second G. 2001. The role of local-level geneflow in enhancing and maintaining genetic diversity. Broadening the genetic base of crop production. CABI Publishing, 452.

Bird A. 2002. DNA Methylation Patterns and Epigenetic Memory. Genes & Development **16**, 6–21.

Blake NK, Lehfeldt BR, Lavin M, Talbert LE. 1999. Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. Genome / National Research Council Canada = Génome / Conseil national de recherches Canada **42**, 351–360.

Bossdorf O, Arcurri D, Richards C, Pigliucci M. 2010. Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology, 541–553.

Bossdorf O, Richards CL, Pigliucci M. 2008. Epigenetics for ecologists. Ecology letters **11**, 106–115.

Bost B, de Vienne D, Hospital F, Moreau L, Dillmann C. 2001. Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. Genetics **157**, 1773–1787.

Brown AHD, Hodgkin T. 2007. Measuring, managing and maintaining crop genetic diversity on farm. Managing biodiversity in agricultural ecosystems, 13–33.

Camacho Villa T, Maxted N, Scholten M, Ford-Lloyd B. 2005. Defining and identifying crop landraces. Plant genetic resources: characterization and utilization **3**, 373–384.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, *et al.* 2012. Biodiversity loss and its impact on humanity. Nature **486**, 59–67.

Ceccarelli S. 1994. Specific adaptation and breeding for marginal conditions. Euphytica **77**, 205–219.

Ceccarelli S. 1996. Adaptation to low high input cultivation. Euphytica **92**, 203–214.

Ceccarelli S. 2009. Evolution, plant breeding and biodiversity. Journal of Agriculture and Environment for International Development **103**, 131–145.

Cervera MT, Ruiz-García L, Martínez-Zapater JM. 2002. Analysis of DNA methylation in Arabidopsis thaliana based on methylation-sensitive AFLP markers. Molecular genetics and genomics: MGG **268**, 543–552.

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, *et al.* 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). The Plant cell **17**, 1033–1045.

Chen T, Li E. 2004. Structure and function of eukaryotic DNA methyltransferases. Current topics in developmental biology **60**, 55–89.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Molecular ecology **21**, 1548–1566.

Crutzen PJ. 2002. Geology of mankind. Nature **415**, 23–23.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401**, 157–161.

Danyluk J, Kane N, Breton G, Limin A, Fowler D, Sarhan F. 2003. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiology **132**, 1849–1860.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Dawson JC, Goldringer I. 2012. Breeding for genetically diverse populations: variety mixtures and evolutionary populations. Organic Crop Breeding, 77–98.

Day T, Bonduriansky R. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. The American naturalist **178**, E18–36.

Dewitt, TJ and Schneider, SM. 2004. Phenotypic variation from single genotypes. In Phenotypic Plasticity: Functional and Conceptual Approaches (ed. T. J. DeWitt and S. M. Scheiner), pp. 1-9. New York: Oxford University Press

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature **418**, 700–707.

Elias M, McKey D, Panaud O, Anstett M, Robert T. 2001. Traditional management of cassava morphological and genetic diversity by the Makushi Amerindians (Guyana, South America): Perspectives for on-farm conservation of crop genetic resources. Euphytica **120**, 143–157.

Elias M, Rival L, Mckey D. 2000. Perception and management of cassava (Manihot esculenta Crantz) diversity among Makushi Amerindians of Guyana (South America). Journal of Ethnobiology **20**, 239–265.

Emperaire L, Peroni N. 2007. Traditional Management of Agrobiodiversity in Brazil: A Case Study of Manioc. Human Ecology **35**, 761–768.

Engelmann F, Engels JMM. 2002. Technologies and strategies for ex situ conservation. Managing plant genetic diversity, 89–103.

Engels J. 2002. *Managing plant genetic diversity [electronic resource]*. CABI.

Enjalbert J, Dawson JC, Paillard S, Rhoné B, Rousselle Y, Thomas M, Goldringer I. 2011. Dynamic management of crop diversity: From an experimental approach to on-farm conservation. Comptes Rendus Biologies **334**, 458–468.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences of the United States of America **95**, 4441–4446.

Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. Benjamin Cummings.

FAO. 1998. *The state of the world's plant genetic resources for food and agriculture.* Rome: FAO.

FAO. 2010. *The second report on the state of the world's plant genetic resources for food and agriculture.* Rome: FAO.

FAO. 2011. www.**fao**.org/

Feldman M, Lupton FGH, Miller TE. 1995. Wheats. In *Evolution of Crops*, 2nd ed., Smartt J and Simmonds NW. Longman Scientific, London.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. Science (New York, N.Y.) **330**, 622–627.

Finckh MR. 2008. Integration of breeding and technology into diversification strategies for disease control in modern agriculture. European Journal of Plant Pathology **121**, 399–409.

Fisher RA. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh **52**, 399–433.

Fisher RA. 1930*a*. *The Genetical Theory Of Natural Selection*. At The Clarendon Press.

Fisher RA. 1930*b*. Genetics, Mathematics, and Natural Selection. Nature **126**, 805–806.

Flatscher R, Frajman B, Schönswetter P, Paun O. 2012. Environmental Heterogeneity and Phenotypic Divergence: Can Heritable Epigenetic Variation Aid Speciation? Genetics Research International **2012**.

Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, Holling CS. 2004. Regime shifts, resislience, and biodiversity in ecosystem management. Annual Review of Ecology, Evolution, and Systematics **35**, 557–581.

Frison EA, Cherfas J, Hodgkin T. 2011. Agricultural Biodiversity Is Essential for a Sustainable Improvement in Food and Nutrition Security. Sustainability **3**, 238–253.

Fu D, Szucs P, Yan L, Helguera M, Skinner JS, von Zitzewitz J, Hayes PM, Dubcovsky J. 2005. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. Molecular Genetics and Genomics: MGG **273**, 54–65.

Gao L, Geng Y, Li B, Chen J, Yang J. 2010. Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. Plant, cell & environment **33**, 1820–1827.

Gepts P, Bettinger R, Brush S, Damania A, Famula T, McGuire P, Qualset C. 2012. Introduction: The Domestication of Plants and Animals: Ten Unanswered Questions. Biodiversity in Agriculture. Cambridge University Press, .

Giménez-Benavides L, García-Camacho R, Iriondo JM, Escudero A. 2011. Selection on flowering time in Mediterranean high-mountain plants under global warming. Evolutionary Ecology **25**, 777–794.

Goldringer I. 2006. Rapid Differentiation of Experimental Populations of Wheat for Heading Time in Response to Local Climatic Conditions. Annals of Botany **98**, 805–817.

Gourdji SM, Mathews KL, Reynolds M, Crossa J, Lobell DB. 2013. An assessment of wheat yield sensitivity and breeding gains in hot environments. Proceedings. Biological sciences / The Royal Society **280**, 20122190.

Gunderson L. 2000. Ecological resilience - in theory and application. Annual Review of Eocology and Systematics **31**, 425–439.

Hall MC, Willis JH. 2006. Divergent Selection on Flowering Time Contributes to Local Adaptation in Mimulus Guttatus Populations. Evolution **60**, 2466–2477.

Hammer K, Teklu Y. 2008. Plant Genetic Resources: Selected Issues from Genetic Erosion to Genetic Engineering. Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS) **109**, 15–50.

Hanocq E, Laperche A, Jaminon O, Lainé A-L, Le Gouis J. 2007. Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **114**, 569–584.

Haudricourt AG. 1964. Nature et culture dans la civilisation de l'igname : l'origine des clones et des clans. L'Homme **4**, 93–104.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S, *et al.* 2007. Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication. Molecular Biology and Evolution **24**, 1506–1517.

Herrera CM, Bazaga P. 2010. Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. The New phytologist **187**, 867–876.

Herrera CM, Bazaga P. 2011. Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. Molecular ecology **20**, 1675–1688.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. Nature **470**, 479–485.

Holliday R. 1994. Epigenetics: An overview. Developmental Genetics **15**, 453–457.

Holt RD, Gomulkiewicz R. 1997. How does immigration influence local adaptation? A reexamination of a familiar paradigm. American Naturalist **149**, 563–572.

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, *et al.* 2005. EFFECTS OF BIODIVERSITY ON ECOSYSTEM FUNCTIONING: A CONSENSUS OF CURRENT KNOWLEDGE. Ecological Monographs **75**, 3–35.

Huang J, Pray C, Rozelle S. 2002. Enhancing the crops to feed the poor. Nature **418**, 678–684.

Jablonka E. 1995. *Epigenetic Inheritance and Evolution: Lamarckian Dimension*. Oxford University Press.

Jablonka E, Lamb MJ. 1989. The inheritance of acquired epigenetic variations. Journal of theoretical biology **139**, 69–83.

Jablonka E, Lamb MJ. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, And Symbolic Variation in the History of Life*. A Bradford Book.

Jablonka E, Raz G. 2009. TRANSGENERATIONAL EPIGENETIC INHERITANCE: PREVALENCE, MECHANISMS, AND IMPLICATIONS FOR THE STUDY OF HEREDITY AND EVOLUTION. QUARTERLY REVIEW OF BIOLOGY **84**, 131–176.

Jackson LE, Pascual U, Hodgkin T. 2007. Utilizing and conserving agrobiodiversity in agricultural landscapes. Agriculture, Ecosystems & Environment **121**, 196–210.

Jarvis DI, Padoch C, Cooper HD. 2007. Biodiversity, agriculture and ecosystem services - Google Scholar. Managing Biodiversity in Agricultural Ecosystems. Jarvis D. I., Padoch C and Copper H. D., 1–12.

Jenuwein T, Allis CD. 2001. Translating the histone code. Science (New York, N.Y.) **293**, 1074–1080.

Johannes F, Colot V, Jansen RC. 2008. Epigenome dynamics: a quantitative genetics perspective. Nature reviews. Genetics **9**, 883–890.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**, 94.

Kamran A, Iqbal M, Navabi A, Randhawa H, Pozniak C, Spaner D. 2013. Earliness per se QTLs and their interaction with the photoperiod insensitive allele Ppd-D1a in the Cutler × AC Barrie spring wheat population. Theoretical and Applied Genetics **126**, 1965–1976.

Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in Arabidopsis thaliana. Annual review of plant biology **55**, 141–172.

Kouzarides T. 2007. Chromatin modifications and their function. Cell **128**, 693–705.

Kremer A, Le Corre V. 2012. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. Heredity **108**, 375–385.

Kuchel H, Hollamby G, Langridge P, Williams K, Jefferies SP. 2006. Identification of genetic loci associated with ear-emergence in bread wheat. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **113**, 1103–1112.

Lachmann M, Jablonka E. 1996. The inheritance of phenotypes: an adaptation to fluctuating environments. Journal of theoretical biology **181**, 1–9.

Lammerts van Bueren E, Backes G, Vriend H, Østergård H. 2010*a*. The role of molecular markers and marker assisted selection in breeding for organic agriculture. Euphytica **175**, 51–64.

Lammerts van Bueren ET, Jones SS, Tamm L, Murphy KM, Myers JR, Leifert C, Messmer MM. 2010*b*. The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. NJAS - Wageningen Journal of Life Sciences **DOI:10.1016/j.njas.2010.04.001**.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics **11**, 204–220.

Lee J-S, Smith E, Shilatifard A. 2010. The Language of Histone Crosstalk. Cell **142**, 682–685.

Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, *et al.* 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC Genomics **13**, 300.

Loskutov IG. 2001. Influence of vernalization and photoperiod to the vegetation period of wild species of oats (Avena spp.). Euphytica **117**, 125–131.

Louette D, Smale M. 2000. Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica **113**, 25–41.

Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J. 2005. Regulation of VRN-1 Vernalization Genes in Normal and Transgenic Polyploid Wheat. Plant Physiol. **138**, 2364–2373.

Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature genetics **38**, 948–952.

Maxted N. 2012. *Agrobiodiversity Conservation*. CABI.

Maxted N, Ford-Lloyd B, Hawkes JG. 1997. Complementary conservation strategies. Plant genetic conservation, the in situ approach. Chapman and Hall, London, 15–39.

Maxted N, Guarino L, Myer L, Chiwona EA. 2002. Towards a methodology for on-farm conservation of plant genetic resources. Genetic Resources and Crop Evolution **49**, 31–46.

Monfreda C, Ramankutty N, Foley JA. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochemical Cycles **22**, n/a–n/a

McKay JK, Latta RG. 2002. Adaptive population divergence: markers, QTL and traits. Trends in Ecology & Evolution **17**, 285–291.

Murphy K, Campbell K, Lyon S, Jones S. 2007. Evidence of varietal adaptation to organic farming systems. Field Crops Research **102**, 172–177.

Nei M. 1972. Genetic distances between populations. Am. Nat. **106**, 283–292.

Olesen JE, Trnka M, Kersebaum KC, Skjelvåg AO, Seguin B, Peltonen-Sainio P, Rossi F, Kozyra J, Micale F. 2011. Impacts and adaptation of European crop production systems to climate change. European Journal of Agronomy **34**, 96–112.

Pál C, Miklós I. 1999. Epigenetic Inheritance, Genetic Assimilation and Speciation. Journal of Theoretical Biology **200**, 19–37.

Parzies HK, Spoor W, Ennos RA. 2000. Genetic diversity of barley landrace accessions (Hordeum vulgare ssp. vulgare) conserved for different lengths of time in ex situ gene banks. Heredity **84 (Pt 4)**, 476–486.

Paun O, Bateman RM, Fay MF, Hedrén M, Civeyrel L, Chase MW. 2010. Stable Epigenetic Effects Impact Adaptation in Allopolyploid Orchids (Dactylorhiza: Orchidaceae). Molecular Biology and Evolution **27**, 2465 – 2473.

Paun O, Bateman RM, Fay MF, Luna JA, Moat J, Hedrén M, Chase MW. 2011. Altered gene expression and ecological divergence in sibling allopolyploids of Dactylorhiza (Orchidaceae). BMC Evolutionary Biology **11**, 113.

Phillips SL, Wolfe MS. 2005. Evolutionary plant breeding for low input systems. JOURNAL OF AGRICULTURAL SCIENCE-CAMBRIDGE- **143**, 245.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. Nature **457**, 843–848.

Raj S, Brautigam K, Hamanishi ET, Wilkins O, Thomas BR, Schroeder W, Mansfield SD, Plant AL, Campbell MM. 2011. Clone history shapes Populus drought responses. Proceedings of the National Academy of Sciences of the United States of America **108**, 12521–12526.

Rao NK. 2004. Plant genetic resources: advancing conservation and use through biotechnology. African Journal of biotechnology **3**, 136–145.

Rapp R, Wendel J. 2005. Epigenetics and plant evolution. NEW PHYTOLOGIST **168**, 81–91.

Raquin A-L, Brabant P, Rhoné B, Balfourier F, Leroy P, Goldringer I. 2008. Soft selective sweep near a gene that increases plant height in wheat. Molecular Ecology **17**, 741–756.

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. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **106**, 895–903.

Rhone B, Remoue 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.

Rhone B, Vitalis R, Goldringer I, Bonnin I. 2010. Evolution of Flowering Time in Experimental Wheat Populations: A Comprehensive Approach to Detect Genetic Signatures of Natural Selection. Evolution **64**, 2110–2125.

Richards EJ. 2009. Quantitative epigenetics: DNA sequence variation need not apply. Genes & development **23**, 1601–1605.

Richards C, Bossdorf O, Verhoeven K. 2010. Understanding natural epigenetic variation. NEW PHYTOLOGIST **187**, 562–564.

Rindos D. 1984. *The origins of agriculture: an evolutionary perspective*. Academic Press.

Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, *et al.* 2009. A safe operating space for humanity. Nature **461**, 472–475.

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, *et al.* 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO Journal **30**, 1928–1938.

Rozenfeld AF, Arnaud-Haond S, Hernández-García E, Eguíluz VM, Serrão EA, Duarte CM. 2008. Network analysis identifies weak and strong links in a metapopulation system. Proceedings of the National Academy of Sciences **105**, 18824–18829.

Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Molecular ecology **14**, 1163–1175.

Sandring S, Riihimäki M-A, 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.

Sasani S, Hemming M, Oliver S, Greenup A, Tavakkol-Afshari R, Mahfoozi S, Poustini K, Sharifi H, Dennis E, Peacock W, *et al.* 2009. The influence of vernalization and daylength on expression of flowering-time genes in the shoot apex and leaves of barley (Hordeum vulgare). Journal of Experimental Botany **60**, 2169–2178.

Smith M, Castillo F, Gomez F. 2001. Participatory plant breeding with maize in Mexico and Honduras. EUPHYTICA **122**, 551–565.

Soengas P, Cartea E, Lema M, Velasco P. 2008. Effect of regeneration procedures on the genetic integrity of Brassica oleracea accessions. Molecular Breeding **23**, 389–395.

Sokal RR, Michener CD. 1958. *A Statistical Method for Evaluating Systematic Relationships*. University of Kansas.

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 / National Research Council Canada = Génome / Conseil national de recherches Canada **43**, 487–494.

Takuno S, Gaut BS. 2013. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proceedings of the National Academy of Sciences of the United States of America **110**, 1797–1802.

Tal O, Kisdi E, Jablonka E. 2010. Epigenetic Contribution to Covariance Between Relatives. Genetics **184**, 1037–1050.

Thomas M. 2011. Gestion dynamique à la ferme de l'agrobiodiversité: relation entre la structure des populations de blé tendre et les pratiques humaines.

Thomas M, Dawson JC, Goldringer I, Bonneuil C. 2011. Seed exchanges, a key to analyze crop diversity dynamics in farmer-led on-farm conservation. Genetic Resources and Crop Evolution **58**, 321–338.

Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics **169**, 1589–1599.

Trevaskis B. 2010. The central role of the VERNALIZATION1 gene in the vernalization response of cereals. Functional Plant Biology **37**, 479–487.

Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES. 2003. MADS box genes control vernalization-induced flowering in cereals. Proceedings of the National Academy of Sciences of the United States of America **100**, 13099–13104.

Trevaskis B, Hemming M, Peacock W, Dennis E. 2006. HvVRN2 responds to daylength, whereas HvVRN1 is regulated by vernalization and developmental status. Plant Physiology **140**, 1397–1405.

Turelli M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theoretical population biology **25**, 138–193.

UNCED. 1992. *Convention on biological diversity*. Department of Public Information, United Nations, New York.

Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, *et al.* 2007. Epigenetic Natural Variation in Arabidopsis thaliana. PLoS Biology **5**, e174.

Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. The New phytologist **185**, 1108–1118.

Via S, Lande R. 1985. Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity. Evolution **39**, 505.

Vijg J, Suh Y. 2013. Genome Instability and Aging. Annual Review of Physiology **75**, 645–668.

Visser ME. 2008. Keeping up with a warming world; assessing the rate of adaptation to climate change. Proceedings of the Royal Society B: Biological Sciences **275**, 649–659.

William GC. 1966. Adaptation and Natural Selection. *Princceton University Press*. Princeton

Wolfe MS. 2000. Crop strength through diversity. Nature **406**, 681–682.

Wolfe MS, Baresel JP, Desclaux D, Goldringer I, Hoad S, Kovacs G, Löschenberger F, Miedaner T, Østergård H, Lammerts van Bueren ET. 2008. Developments in breeding cereals for organic agriculture. Euphytica **163**, 323–346.

Wood D, Lenné JM. 1997. The conservation of agrobiodiversity on-farm: questioning the emerging paradigm. Biodiversity and Conservation **6**, 109–129.

Wood D, Lenné JM. 1999. *Agrobiodiversity: characterization, utilization, and management*. CABI Pub.

Worland T, Snape JW. 2001. Genetic basis of worldwide wheat varietal improvement. In *The World Wheat Book, A History of Wheat Breeding*, ed., Bonjean AP, Angus WJ, Lavoisier Publishing Paris.

WRI, IUCN, UNEP. 1992. Global biodiversity strategy : guidelines for action to save, study, and use earth's biotic wealth sustainably and equitably. **WRI, IUCN, UNEP**

Lev-Yadun S, Gopher A, Abbo S. 2000. The Cradle of Agriculture. Science **288**, 1602–1603.

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J. 2006. The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proceedings of the National Academy of Sciences **103**, 19581–19586.

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. 2004. The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. Science **303**, 1640–1644.

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 of the United States of America **100**, 6263–6268.

Yi C, Zhang S, Liu X, Bui HT, Hong Y. 2010. Does epigenetic polymorphism contribute to phenotypic variances in Jatropha curcas L.? BMC Plant Biology **10**, 259.

Zhang Y-Y, Fischer M, Colot V, Bossdorf O. 2013. Epigenetic variation creates potential for evolution of plant phenotypic plasticity. The New phytologist **197**, 314–322.

Zhang M, Kimatu JN, Xu K, Liu B. 2010. DNA cytosine methylation in plant development. Journal of Genetics and Genomics = Yi Chuan Xue Bao **37**, 1–12.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, *et al.* 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell **126**, 1189–1201.

Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, Fan J, Yang S, Hu L, Leung H, *et al.* 2000. Genetic diversity and disease control in rice. Nature **406**, 718–722.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2006. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics **39**, 61–69.

ANNEX

At the start of my PhD, before choosing KASPAR method (Kbioscience) of genotyping due to its cost and time effectiveness, genotyping was initially planned to be done at Le Moulon for candidate genes and at genotyping platform of INRA at Clermont-Ferrand for SSR neutral markers. As *VRN-A1* promoter alleles were among those to be genotyped, I worked, to improve the existing markers for these alleles. One of the alleles, *Vrn-A1a,* was reported to have a duplication and a 91bp deletion in one of the duplicates, therefore presenting a double band pattern on the gel. Interestingly, it was observed that the intensity of the second band was inconsistent. This raised suspicions that PCR artefact could be at the origin of this double band pattern in *Vrn-A1a* promoter amplification. So a series of experiments was carried out to test this hypothesis and the findings of this work are presented in the following pages in the form short scientific note.

Analysing *VRN-A1* promoter duplications in wheat: a case of PCR artefact

Introduction

Unravelling genetic determinism of complex traits is of major academic and applied interest. One strategy relies on the extensive description of allelic variations at QTLs, i.e. building portfolios describing molecular variations at given loci, along with their expected effects on quantitative traits. If new sequencing methods facilitates an extensive access to single base DNA variations as well as copy-number variations, most of these techniques rely on a preliminary PCR amplification of genomic DNA, a reliable step, but sometimes prone to artefacts, especially on complex, repetitive DNAs. We report here a PCR artefact due to presence of short sequences in direct repeat, interpreted so far as a duplication at *VRN-A1* promoter, a region of major effect on wheat flowering date.

Transgenic and mutant analysis data shows that *VRN1* is a flowering promoter, with a critical role in the floral transition pathway of wheat (Loukoianov *et al.*, 2005; Shitsukawa *et al.*, 2007). In the hexaploid genome of T. aestivum, the dominant-spring type allele in genome A (Vrn-1A), has a stronger effect than the dominant-spring type alleles in genome B and D (*Vrn-1B, Vrn-1D,* (Pugsley, 1971; Trevaskis *et al.*, 2003). Yan *et al.*, (2004) developed genome specific primers for the promoter regions of *VRN1*. Reporting allelic variations for the *VRN-A1* copy, these authors described a *Vrn-A1a* allele, presenting a double band pattern, contrasting with the single band *vrn-1A* allele (recessive/winter allele). Sequence analysis showed that the two fragments differed from the recessives *vrn-1A* allele by the insertion of a foldback element (222-bp), partly truncated (131-pb) in the smaller fragment. The double band pattern, as well as the presence of a SNP, leads the author to hypothesize the presence of a duplication in the promoter region of *Vrn-A1a* allele. Working with the same primer pair and PCR conditions than (Yan *et al.*, 2004), we observed that the amplification of the *Vrn-A1a* allele always produced fluctuating results (Figure 1, lane 2-3), with a lower band showing poor or no signal. Such low intensity of the shortest fragment, also observable in other studies (see fig.2 in Nowak and Kowalczyk, (2010) or in Zhang *et al.*, (2008)), is not in agreement with the expected outcome of classic PCR competition. We therefore have sought for structural specificities of the targeted alleles. Indeed, *VRN-A1-a* contains a foldback element (Yan *et al.*, 2004), presenting two short sequences in direct repeats, and nested position (Figure 2). As the short fragment of allele 1 correspond to the deletion of the inner repeat (91 bp), such lower band could be produced by recombination during the PCR amplification. DNA recombination and chimera production during PCR has been effectively reported, generally occurring on fragments presenting strong internal homology, like transposons LTR. One hypothesis concerning their origins is that incompletely extended fragments can serve as primers, and anneal to closely related sequences generating recombinations (Meyerhans *et al.*, 1990)

On the basis of these observations and bibliographic background, we assessed whether PCR artefact could be at the origin of a double band pattern in *Vrn-A1a* promoter amplification.

Material and methods

Our strategy relies on the cloning of Vrn-A1 promoter, and subsequent PCR study. We first selected, within our collection of genotypes possessing the *Vrn-A1a* allele, two distinct lines presenting the clearest double-

band patterns (Figure 1). The two lines G1 & G2 were taken from a dynamic management program of wheat genetic resources, and have been previously characterized as carrying the *Vrn-A1a* allele (Rhone *et al.*, 2008). For *VRN-A1* promoter amplification, we used the genome A specific primers developed by (Yan *et al.*, 2004) (*VRN1*AF: GAAAGGAAAAATTCTGCTCG; *VRN1*R: TGCACCTTCCCSCGCCCCAT). All PCR reactions were performed in 22 µl. The reference PCR mix contained 1X buffer (16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl and 0.01% TWEEN-20), 1.5 mM $MgCl_2$, 0.25 mM dNTP, 0.5 µM of forward and reverse primers, 1 Unit of Taq polymerase and 10 ng of genomic DNA. The reference PCR program started with denaturing at 95°C for 5 minutes, then 10 cycles of touch down (94°C x 1 mn ; 65°C x 1 mn, with a 1°C decrease at each cycle; 72°C x 1 mn 30 sec, followed by 21 cycles (94°C x 1 mn ; 55°C x 1mn ; 72°C x 1 mn 30 sec). Effect of variations around this protocol on relative intensities of the two amplified bands was also assessed (see legend of Figure 1).

PCR products of the two selected genotypes were used directly for cloning in a pGEM-T Vector system I (Promega). Two additional spring genotypes, deleted for the foldback element, were also included as cloning controls (data not shown). After selection of transformed colonies on LB medium + ampicillin + Xgal, and PCR check of insert presence, plasmid DNA was extracted for 12 colonies per genotype using QIAGEN Plasmid Kit. Insert size was analysed after an Eco52I digestion. A subset of 8 colonies per genotype was sequenced using plasmid primer T7PROM. A direct PCR of *VRN-A1*prom was performed and compared to insert extraction (Figure 3).

To assess the stability of double band patterns, we selected two colonies per genotype presenting a single high band, in order to purify insert producing only the highest band. The selected colonies were re-plated in a Petri-dish with LB medium + ampicillin + Xgal, and 16 colonies were isolated and *VRN-A1*prom bands were re-checked by direct PCR on colonies. A second round of this high-band colony selection, re-plating, and PCR check was performed (figure 3).

Results:

Important variations in the relative intensity of the two *VRN-1Aa* bands were observed according to changes in the PCR protocol, with an overall higher intensity of the higher band (Figure 1). Increase in elongation temperature, decrease in elongation time or increase in number of cycles reduced the intensity of the lower band, while use of Q-solution (PCR quality enhancer) slightly increased it. Even repetitions of the same PCR at two different dates displayed variation in band intensities (data not shown), stressing the unsteady nature of the lower band.

When attempting to isolate through cloning the two bands of the *VRN-1Aa* PCR product, a surprising persistence of the two bands was detected by PCR or plasmid purification in most of the clones produced, in two independent cloning experiments involving two different *VRN-A1a* lines. If lower band was sometimes isolated in a clone, higher band alone was almost never found. Sequencing confirmed that the two bands were the two expected fragments of the *Vrn-A1a* allele. When trying to further isolate the two bands, two successive rounds of sub-cloning produced the same results: higher band colonies produced either double band or single lower band colonies, while lower bands were only producing lower band colonies. These results show the ability of the long *Vrn-A1a* fragment to produce the truncated band, during both PCR amplification or plasmid replication.

Discussion:

Due to the poor reproducibility of the two band pattern observed when amplifying by PCR the promoter of *VRN-A1a* allele, we attempted to isolates the two PCR products. While shortest band was easily cloned, we failed to steadily isolate the long band (700pb), as the many clones studied were still showing a double band pattern, both after amplification or plasmid purification. Further attempts of isolate the higher band through sub-clone selection failed as well. These results clearly demonstrate that the presence of the complete foldback element allows the neo-formation of the short band, during PCR reaction, as well as during plasmid replication. Recombination within the foldback element is the most likely explanation to these observations. Effectively, the foldback element presents direct repeats (figure 1), which are known to be at the origin of PCR recombination. The most admitted explanation is that incomplete PCR products, ending on the first repeat of the fragment, can serve as primers in the next PCR step (Meyerhans *et al.*, 1990; Judo *et al.*, 1998). These incomplete fragments can mispair on the second repeat, and produce fragments deleted of the inner sequence (131 bp). Direct repeats can also favour deletions during plasmid replication (Sumegi *et al.*, 1997) or viral replication (Kong and Masker, 1994) in bacteria, and also affect other organisms (yeast: Phadnis *et al.*, 2005; mice: Würtele *et al.*, 2005, wheat: Ogihara *et al.*, 1988. Such direct repeats might favour deletions or rearrangements through the secondary structures they promote (Bowater and Wells, 2001).

Our result therefore stresses the importance of considering PCR aftefacts when studying genomic regions presenting direct (or inverted) repeats. For the specific case of *VRN-A1*, this result questions the reality of the duplication of the promoter in genotypes carrying *VRN-A1a* allele. As a SNP is found in the Yan *et al.*, (2004) study, the presence of a duplication of the promoter in the lines studied is not questioned by the present study, even if sequencing process can be sometimes prone to recurrent errors (see for example Zaranek *et al.*, 2010). In accordance with (Yan *et al.*, 2004), other duplications of *VRN-A1* promoter have been described in wheat germplasm by Golovnina *et al.*, (2010). In a more specific genetic analysis, the presence of full-length, functional duplication (or triplication) in hexaploïd germplasm has been demonstrated. As for promoter duplication, presence of tandem duplication of VRN-A1 was assumed from observed CNV segregations in biparental progenies. Interestingly, vernalisation requirement was correlated to the copy number variations (CNV), confirming the importance of these structural changes for functional variation at adaptive traits (Gokcumen *et al.*, 2011).

More generally, in the context of the development of high-throughput sequencing, the known risks of PCR recombination and chimeric sequence generation in a highly duplicated genome as hexaploid wheat should motivate extra-care in the DNA preparation prior sequencing (Lenz and Becker, 2008), as well as in the data analysis.

Reference:

Bowater RP, Wells RD. 2001. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. Progress in Nucleic Acid Research and Molecular Biology **66**, 159–202.

Gokcumen O, Babb PL, Iskow RC, Zhu Q, Shi X, Mills RE, Ionita-Laza I, Vallender EJ, Clark AG, Johnson WE, *et al.* 2011. Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. Genome Biology **12**, R52.

Golovnina KA, Kondratenko EY, Blinov AG, Goncharov NP. 2010. Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. BMC Plant Biology **10**, 168.

Judo MSB, Wedel AB, Wilson C. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Research **26**, 1819 –1825.

Kong D, Masker W. 1994. Deletion between direct repeats in T7 DNA stimulated by double-strand breaks. Journal of bacteriology **176**, 5904–5911.

Lenz TL, Becker S. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis RID D-1387-2009. Gene **427**, 117–123.

Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J. 2005. Regulation of VRN-1 Vernalization Genes in Normal and Transgenic Polyploid Wheat. Plant Physiol. **138**, 2364–2373.

Meyerhans A, Vartanian JP, Wain-Hobson S. 1990. DNA recombination during PCR. Nucleic Acids Research **18**, 1687–1691.

Nowak M, Kowalczyk K. 2010. Allelic Variation at the VRN-1 Locus of Polish Cultivars of Common Wheat (Triticum aestivum L.). Acta Biologica Cracoviensia Series Botanica **52**, 86–91.

Ogihara Y, Terachi T, Sasakuma T. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proceedings of the National Academy of Sciences **85**, 8573 – 8577.

Phadnis N, Sia R, Sia E. 2005. Analysis of repeat-mediated deletions in the mitochondrial genome of Saccharomyces cerevisiae. Genetics **171**, 1549–1559.

Pugsley A. 1971. A genetic analysis of the spring-winter habit of growth in wheat. Aust. J. Agric. Res. **22**, 21–31.

Rhone B, Remoue 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.

Shitsukawa N, Ikari C, Shimada S, Kitagawa S, Sakamoto K, Saito H, Ryuto H, Fukunishi N, Abe T, Takumi S, *et al.* 2007. The einkorn wheat (Triticum monococcum) mutant, maintained vegetative phase, is caused by a deletion in the VRN1 gene. Genes & Genetic Systems **82**, 167–170.

Sumegi A, Birko Z, Szeszak F, Vitalis S, Biro S. 1997. A short GC-rich sequence involved in deletion formation of cloned DNA in E-coli. Acta Biologica Hungarica **48**, 275–279.

Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES. 2003. MADS box genes control vernalization-induced flowering in cereals. Proceedings of the National Academy of Sciences of the United States of America **100**, 13099–13104.

WRI, IUCN, UNEP. 1992. Global biodiversity strategy : guidelines for action to save, study, and use earth's biotic wealth sustainably and equitably.

Würtele H, Gusew N, Lussier R, Chartrand P. 2005. Characterization of in vivo recombination activities in the mouse embryo. Molecular Genetics and Genomics **273**, 252–263.

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.

Zaranek AW, Levanon EY, Zecharia T, Clegg T, Church GM. 2010. A Survey of Genomic Traces Reveals a Common Sequencing Error, RNA Editing, and DNA Editing. PLoS Genet **6**, e1000954.

Zhang XK, Xiao YG, Zhang Y, Xia XC, Dubcovsky J, He ZH. 2008. Allelic Variation at the Vernalization Genes , , , and in Chinese Wheat Cultivars and Their Association with Growth Habit. Crop Science **48**, 458–470.

FIGURES:

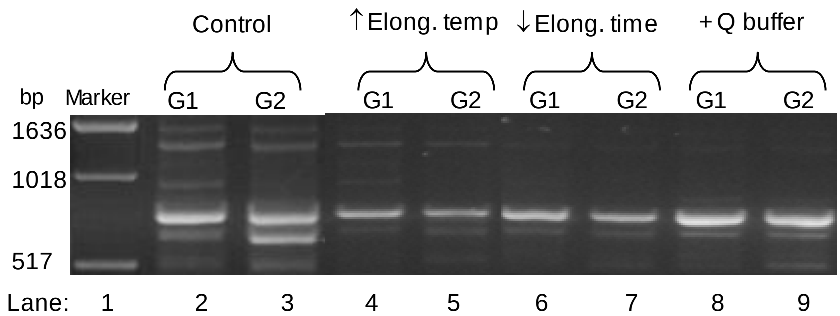


Figure 1: PCR amplification of VRN-1A promoter region with primers *VRN1*-AF/*VRN1*-R

Modifications in reagent concentrations and PCR program were done to test their effect on band intensities, as compared to control PCR on two genotypes carrying *Vrn-1Aa* allele (lane 2-3): a) increasing the elongation temperature to 72°C instead of 65°C (lane 4-5), b) decreasing the elongation time to 20 seconds instead of 1m 30sec (lane 6-7) and c) including 1 U of Q buffer from Qiagene in PCR mix (enhances PCR quality: lane 8-9).

```
                   10        20        30        40        50        60        70        80        90
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Vrn-A1a(HB) TTAAAAACCGGAAAAAAAATTATATGAGACCAGGTCTCATATAAATCAGGTGAGACCCGCCCTGATGAATGACATGTGGCATTCACAAATC
Vrn-A1a(LB) TTAAAAACCGGAAAAAAAATTATATGAGACCAGGTCTCATATAAATCAGGTGAGACCCGCCCTGATGAATGACA~~~~~~~~~~~~~~~~~

                  100       110       120       130       140       150       160       170       180
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Vrn-A1a(HB) ACAAAGCATCTAATCTCTCCCCCCCTGATTTCAGGTGGGGGTGGGGTGGATGCTTTGTGATTTGTGAATGACACGTGTCATCCATCAGG
Vrn-A1a(LB) ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CGTGTCATCCATCAGG

                  190       200       210       220       230       240
           ....|....|....|....|....|....|....|....|....|....|....|....|
Vrn-A1a(HB) ATGGGTCTCACCTGCTAATCCGTGAGACCTGGTCTCATAGAATTTTTTTCCTTAAAAACC
Vrn-A1a(LB) ATGGGTCTCACCTGCTAATCCGTGAGACCTGGTCTCATAGAATTTTTTTCCTTAAAAACC
```

Figure 2: Sequence of foldback elements with direct repeat represented in bold. HB denotes higher band and LB denotes lower band

Figure 3: Pictoral representation of cloning experiment and gel photos after each Direct PCR (Colony PCR)