



**HAL**  
open science

## Diversifying mechanisms in the on-farm evolution of crop mixtures

Mathieu Thomas, Stéphanie Thépot, Nathalie Galic, Sophie Pin, Carine  
Remoue, Isabelle Goldringer

► **To cite this version:**

Mathieu Thomas, Stéphanie Thépot, Nathalie Galic, Sophie Pin, Carine Remoue, et al.. Diversifying mechanisms in the on-farm evolution of crop mixtures. *Molecular Ecology*, 2015, 24 (12), pp.2937-2954. 10.1111/mec.13214 . hal-01535269

**HAL Id: hal-01535269**

**<https://hal.science/hal-01535269v1>**

Submitted on 16 May 2018

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

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

# Diversifying mechanisms in the on-farm evolution of crop mixtures

MATHIEU THOMAS,\* STÉPHANIE THÉPOT,† NATHALIE GALIC,\* SOPHIE JOUANNE-PIN,\*  
CARINE REMOUÉ\* and ISABELLE GOLDRINGER\*

\*INRA, UMR 0320/UMR 8120 *Génétique Quantitative et Évolution - Le Moulon, F-91190 Gif-sur-Yvette, France*, †ARVALIS  
*Institut du Végétal, route de Chateaufort, F-91190 Villiers-Le-Bacle, France*

## Abstract

While modern agriculture relies on genetic homogeneity, diversifying practices associated with seed exchange and seed recycling may allow crops to adapt to their environment. This socio-genetic model is an original experimental evolution design referred to as on-farm dynamic management of crop diversity. Investigating such model can help in understanding how evolutionary mechanisms shape crop diversity submitted to diverse agro-environments. We studied a French farmer-led initiative where a mixture of four wheat landraces called 'M'élange de Touselles' (MDT) was created and circulated within a farmers' network. The 15 sampled MDT subpopulations were simultaneously submitted to diverse environments (e.g. altitude, rainfall) and diverse farmers' practices (e.g. field size, sowing and harvesting date). Twenty-one space-time samples of 80 individuals each were genotyped using 17 microsatellite markers and characterized for their heading date in a 'common-garden' experiment. Gene polymorphism was studied using four markers located in earliness genes. An original network-based approach was developed to depict the particular and complex genetic structure of the landraces composing the mixture. Rapid differentiation among populations within the mixture was detected, larger at the phenotypic and gene levels than at the neutral genetic level, indicating potential divergent selection. We identified two interacting selection processes: variation in the mixture component frequencies, and evolution of within-variety diversity, that shaped the standing variability available within the mixture. These results confirmed that diversifying practices and environments maintain genetic diversity and allow for crop evolution in the context of global change. Including concrete measurements of farmers' practices is critical to disentangle crop evolution processes.

*Keywords:* crop biodiversity, local adaptation, on-farm dynamic management, on-farm evolutionary experiment

## Introduction

Genetic diversity is assumed to be of major importance for the adaptation of both wild and cultivated species to future environmental changes (Barrett & Schluter 2008; Mercer & Perales 2010). As farmers around the world use various farming practices and grow different species for different uses in different agroecosystems, they *de facto* contribute to the on-farm dynamic manage-

ment of crop diversity. This is a complementary strategy to *ex situ* conservation, which allows genetic resources to continuously adapt to changing environments (Bretting & Duvick 1997; Maxted *et al.* 1997; Negri & Tiranti 2010; Enjalbert *et al.* 2011). While in dynamic management, crop populations are mainly submitted to natural selection in experimental conditions, in the context of on-farm dynamic management systems, crop populations are submitted to both natural selection and human-mediated selective pressures through different farming practices (Enjalbert *et al.* 2011). Several studies showed that some farming prac-

Correspondence: Mathieu Thomas, Fax: +331 693 323 40;  
E-mail: mateothom@gmail.com

tices maintain a high level of genetic and phenotypic diversity as shown for the management of cassava in the Makushi Amerindian community in Guyana where farmers use and multiply volunteer plants in addition to their clonal varieties (Elias *et al.* 2001). Moreover, Mexican farmers' practices have also been shown to induce assortative mating in maize populations, leading to a spacial distribution of the phenotypic diversity and maintaining a relatively high level of neutral diversity by seed circulation (Pressoir & Berthaud 2003a). This particular practice of seed circulation is mediated by social organization which can strongly reshapes crop genetic diversity, as it was shown for wheat in France (Thomas *et al.* 2012). In particular, the seed circulation is often driven by specific cultural rules such as those organizing the weddings as observed for cassava in Gabon (Delêtre *et al.* 2011) and for sorghum in Kenya (Labeyrie *et al.* 2014). Most seed sources usually come from the same community, so long-distance seed circulation rarely occurs (Louette *et al.* 1997; Bellon *et al.* 2011; Thomas *et al.* 2011; Samberg *et al.* 2013). Another common farming practice consists in mixing seeds of several varieties and then resowing the mixture of the harvested seeds. Such mixtures of varieties are relatively widespread due to their year in, year out robustness that enables them to tolerate variations in biotic and abiotic pressures (Dawson & Goldringer 2012). In a mixture of genotypes, adaptation can result from an increase in frequency of the most adapted component or from the emergence of a new genotype with higher fitness through recombination, depending on the mating system. In a mixture, individual fitness is expected to depend on the interrelated effects of local adaptation and competitive ability (Zuppinge-Dingley *et al.* 2014). Yet, in a mixture of landraces, that is genetically heterogeneous populations, selection might occur within and among components and makes the mixture even more complex. But up to now, little attention has been paid to the genetic mechanisms that underlie the micro-evolution of such populations simultaneously submitted to seed circulation, mixture and natural/human-mediated selection. Such subdivided populations are simultaneously submitted to natural selection and human-mediated selection by farmers, in addition to genetic drift and migration.

In this study, we studied a recent mixture called 'Mélange de Touselles' (MDT) composed of three landraces of bread wheat (*Triticum aestivum*) and one landrace of cone wheat (*Triticum turgidum* subsp. *turgidum*). This mixture was created by one farmer in 2001 and has been continuously cultivated on his farm since then. It was distributed by this farmer to other farmers from 2004 onward. Therefore, a set of MDT populations has been evolving within a group of farmers located in different

sites in France, but also in Italy and in the Netherlands. In this study, we considered that each seed circulation event corresponded to a founding effect in order to account for the underlying social organization. To our knowledge, this original farmer-led design can be considered as the first ever on-farm evolutionary experiment.

In this context of an evolutionary experiment, the main goal was (i) to characterize the genetic structure and the spatial and temporal differentiation pattern of a self-pollinated crop mixture recently introduced in different environments and (ii) to identify some of the underlying evolutionary mechanisms. The evolutionary dynamics of the mixture was studied using on polymorphisms at neutral markers and at genes associated with flowering time variation. We focused on earliness as it is an important adaptive trait involved in the synchronization of the plant cycle with the environment. A combined approach relying on a discriminant analysis and a dedicated network-based method was used to decipher the complex genetic structure of the mixture.

A temporal and spatial sampling of MDT populations associated with genetic and phenotypic analysis was carried out and allowed us to (i) describe the general genetic structure of the landraces composing the mixture; (ii) estimate the differentiation of the mixture at the phenotypic, neutral genetic and gene levels and (iii) understand the genetic and evolutionary mechanisms that underlie the differentiation and response to selection of the different populations. Last of all, this work aimed at understanding the combined role of farmers' practices and environment on the rapid evolution of a crop mixture and it highlights the crucial role played by within-population variability in adaptation.

## Materials and methods

### *Description of the on-farm evolutionary experiment*

The mixture was developed by a farmer (called farmer A), who decided to re-establish four local landraces of the southeastern region of France at the end of the 1990s. Landraces historically grown in that region disappeared in the middle of the twentieth century when modern agriculture replaced landraces and old varieties by elite material. He obtained around 50 seeds of each of the four varieties: *Touselle Anone* (TAN), *Touselle Blanche Barbue* (TBB), *Touselle Blanche de Provence* (TBP) and *Touselle sans Barbe* (TSB). It is important to note that TBB is an allotetraploid variety of wheat (*Triticum turgidum* subsp. *turgidum*), whereas the others are allohexaploid (*Triticum aestivum*) from the national French GenBank of Clermont-Ferrand. The landraces were grown in small plots from 1997 to 2001. This allowed the farmer A to increase seed number and therefore

plot size from 1 m<sup>2</sup> in 1997–1998 to 10 m<sup>2</sup> in 1998–1999, 100 m<sup>2</sup> in 1999–2000 and 1000 m<sup>2</sup> in 2000–2001, respectively. In 2001, bad weather conditions caused important lodging (plants bent down to the ground level before being harvested) and farmer A decided to harvest the four varieties together. This mixture was called *Mélange de Touselles* (MDT). It is composed of an unknown initial proportion of *T. aestivum* and *T. turgidum*, and it has been maintained since then in a large plot of around one hectare. In 2004, farmer A started to distribute seed lots of the mixture to other farmers. We studied temporal samples of farmer A, as well as samples of mixtures from 13 other farmers who had been given MDT in 2004, 2005 and 2006 (see Table 1). Further spreading of the MDT mixture will not be considered in this study. This collective experience was self-organized by farmers. It was studied as a case of on-farm experimental evolution of crop populations.

#### Data collection

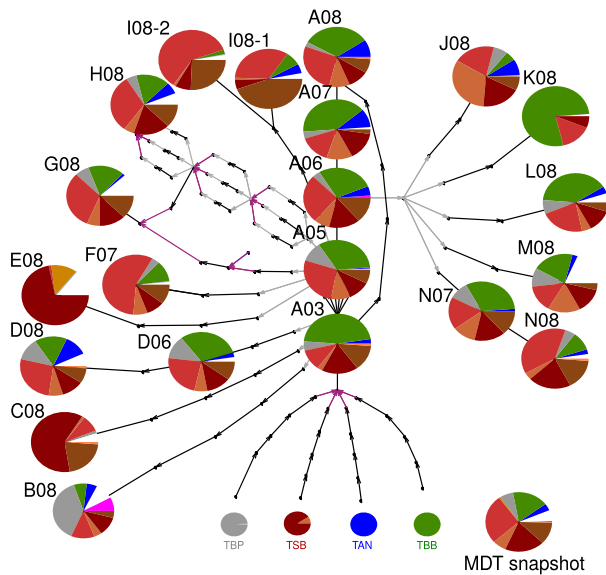
All the information collected during interviews received the prior informed consent of the interviewed people. Information about MDT circulation was obtained

through several interviews with farmer A. Directed telephone interviews were carried out with 10 French farmers who grew the mixture to collect information about their farming practices such as the plot size, sowing density, the previous crop grown in the field, the sowing date, the weed control practices, the harvest date. Similar information about the two populations grown in the Netherlands and the other one in Italy was obtained from the Farm Seed Opportunity project during which the mixture was distributed to Dutch and Italian farmers (Serpoly *et al.* 2011). Based on this information, we built a seed circulation and multiplication network (Fig. 1). Demographic size of each population was estimated using the sowing density and the plot size information. In addition, the curator in charge of the wheat collection at the national GenBank was also interviewed to gather information about multiplication and growing practices during seed multiplication and regeneration of the accessions. Specific information about management practices of TAN, TBB, TBP and TSB was also collected.

*Sampling strategy.* To depict the evolution of the recent MDT mixture, 21 samples were collected from 14

**Table 1** Description of the *Mélange De Touselles* (MDT) samples (Origin: character to identify the different people who provided MDT seed lot samples, Harvest year: year of the last harvest before studying the seed lot sample, Plot size: 1 = ranging from 1 to 10 m<sup>2</sup>, 2 = ranging from 10 to 100 m<sup>2</sup>, 3 = larger than 100 m<sup>2</sup>; #G: Number of generations as the MDT was created by farmer A, NBGENOUT: number of generations grown outside A’s farm), Long.: longitude, Lat.: latitude, Alt.: altitude)

Sample name	Origin	Harvest year	Plot size	#G	NBGENOUT	Long.	Lat.	Alt. (m)
A03	A	2003	3	2	0	4.340	43.790	109
A05	A	2005	3	4	0	4.340	43.790	109
A06	A	2006	3	5	0	4.340	43.790	109
A07	A	2007	3	6	0	4.340	43.790	109
A08	A	2008	3	7	0	4.340	43.790	109
B08	B	2008	3	7	4	4.030	44.020	131
C08	C	2008	2	7	4	5.370	44.900	1068
D06	D	2006	3	7	2	-0.230	47.460	21
D08	D	2008	3	7	4	-0.230	47.460	21
E08	E	2008	3	7	3	5.400	45.530	495
F07	F	2007	1	6	2	6.310	46.170	500
G08	G	2008	2	7	3	-1.130	47.010	90
H08	H	2008	1	7	3	0.650	46.670	33
I08-1	I	2008	1	7	3	5.270	47.560	296
I08-2	I	2008	2	7	3	5.270	47.560	296
J08	J	2008	1	7	2	11.430	45.620	260
K08	K	2008	1	7	2	1.710	49.120	127
L08	L	2008	1	7	2	5.230	52.180	-3
M08	M	2008	1	7	2	13.900	42.020	577
N07	N	2007	1	6	1	-0.650	47.420	52
N08	N	2008	1	7	2	-0.650	47.420	52
GBTAN	GenBank	2004	1	—	—	—	—	—
GBTBB	GenBank	2003	1	—	—	—	—	—
GBTBP	GenBank	2004	1	—	—	—	—	—
GBTSB	GenBank	2003	1	—	—	—	—	—



**Fig. 1** Frequency of the different genetic groups among the sampled populations of *Mélange De Touselles* (MDT). At the bottom, the genetic composition of the four reference landraces: *Touselle Anone* (TAN), *Touselle Blanche Barbe* (TBB), *Touselle Blanche de Provence* (TBP) and *Touselle sans Barbe* (TSB). The five vertically aligned pies are time series of farmer A's population. The other pies correspond to populations grown by different farmers for one to four generations. TBP is in grey, TSB is divided into five subgroups (TSB1: red, TSB2: light red, TSB3: dark red, TSB4: brown and TSB5: salmon pink), TAN is in blue, TBB in green, *Florence-Aurore* (FLA) in pink, *Touselle des Hautes-Alpes* (THA) in orange and unassigned in white. Multiplication events are represented by black arrows, diffusion by grey arrows and mixture by purple arrows. Pies size is proportional to the number of individuals: 32 plants per reference landrace sample and almost 80 plants per population sample.

farmers including farmer A (Fig. S1 in Supporting information). In the dynamics of the MDT populations, we considered that a new subpopulation appeared when a seed lot was sown on a new farm, or on the same farm with different farming practices (e.g. when farmer I chose to grow MDT in distinct plots of different sizes: I08-1 and I08-2). The labels of the samples summarize (i) the seed lot owner (first letter); (ii) the harvest year (first two digits); and (iii) in some cases, information relative to plot size (last digit) (See Table 1 for more details).

The material included five samples corresponding to a time series from farmer A's population (A03, A05, A06, A07, A08, respectively), two temporal samples from farmer D's population (D06, D08) and from farmer N's population (N07, N08). All other samples corresponded to populations from different farms where the mixture was grown for 1–4 years. A snapshot of the different versions of MDT populations was taken from the subset of fourteen populations sampled in 2008: A08,

B08, C08, D08, E08, G08, H08, I08-1, I08-2, J08, K08, L08, M08 and N08. The F07 sample was also included in the snapshot as farmer F did not grow MDT in 2008, leading to a sample of 15 populations. In addition, samples of the four varieties that were used to build the mixture were provided by the national French GenBank as a reference for the initial composition of the MDT mixture (GBTAN, GBTBB, GBTBP and GBTBSB).

**Molecular analyses.** In December 2008, around hundred seeds from the 21 MDT samples each and around fourteen seeds from the four GenBank varieties were sown in individual pots in the glasshouse at Le Moulon experimental station. In January 2009, leaf samples were taken from 80 healthy plants randomly chosen per population, while only 32 individuals were sampled for each of the four varieties, that is 1808 individuals altogether. For each plant, total DNA was extracted from 100 mg of fresh material following a protocol derived from the QIAGEN DNeasy 96 Plant Kit (QIAGEN, Basel, Switzerland). Fourteen microsatellite markers (single sequences of tandem repeats) developed by Röder *et al.* (1998), one (wmc231) by Somers *et al.* (2004) and a bilocus marker (cfd71) developed by Guyomarc'h *et al.* (2002) were used to genotype the 1808 plants. This set of 17 markers covered 17 of the 21 chromosomes of bread wheat. Only chromosomes 1A, 6A, 6B and 7D were not covered. PCR protocols were adapted from Röder *et al.* (1998) and Guyomarc'h *et al.* (2002). Amplified fragments were separated on an ABI 3130xl semi-automatic sequencer (Applied Biosystems) and analysed with GENEMAPPER 3.7 (Applied Biosystems). Individuals with more than 6 missing data among the 17 microsatellite markers were discarded. After genotyping, the final data set contained 1793 individuals of the initial 1808. Sometimes bands were detected for markers mapped on the D genome in individuals from the initial GBTBB sample (allotetraploid genome AB) although microsatellite markers were designed to be specific for each genome (A, B and D, respectively). To avoid overestimating interspecific crosses, bands that were specific to the initial GBTBB landrace were noted among the 17 markers to be used as a reference. Then, a screening was performed to detect and to separate TBB individuals from the whole data set based on these signature bands (407 individuals). These TBB individuals were scored with a specific allele code for loci mapped in the D genome and were studied independently in some analyses. We further assumed that all TBBs identified in MDT populations were allotetraploid AB, while the rest was allohexaploid ABD.

Four polymorphisms located in candidate genes associated with earliness (Rousset *et al.* 2011) were genotyped to detect whether these genes were submitted to

selection during the evolution of the mixture. One single nucleotide polymorphism (SNP; C/T) located in the first intron of the FTA gene (Bonnin *et al.* 2008) was genotyped, as well as three polymorphisms located in two of the three copies of the *VRN-1* gene (A and D genomes):

- one polymorphism by duplication, insertion and deletion in the promoter region of *VRN-1A* (named *VRN-1A<sub>prom</sub>*), identified by Yan *et al.* (2004),
- one substitution in the seventh exon of *VRN-1A* (*VRN-1A<sub>ex7</sub>*), detected by Sherman *et al.* (2004),
- a 4-kb deletion in the first intron of *VRN-1D* (*VRN-1D*), detected by Fu *et al.* (2005)

*Glasshouse experiment.* In February 2009, all the plants sampled for genotyping were transplanted to ground soil under a plastic tunnel to be submitted to natural vernalization at the 4-leaf stage following a completely randomized two-block design. Distance between plants within rows was 10 cm and it was 20 cm between rows. Viability, plant size and heading date as a proxy for flowering time were scored at the individual level during plant development. An individual was considered as having headed when half of the ear was out of the leaf sheath. It is important to note that 25% of the plants measured more than 1.60 m in the experimental conditions. This may have caused competition among neighbouring plants.

#### Data analysis

The evolution of diversity in the Mélange de Touselles was studied by comparing:

- the four reference varieties (GBTAN, GBTBB, GBTBP and GBTB) pooled together with equal frequency, named virtual MDT as the initial frequencies of the four components were not known,
- the first mixture available, A03, used as initial reference population,
- the MDT snapshot composed of the fifteen populations sampled in 2007 or 2008.

*Detection of the genetic structure.* The aim of the genetic structure detection was to identify the main genetic groups and to characterize the level of the within-group genetic diversity. As wheat is highly autogamous, all the loci we studied were in strong linkage disequilibrium (data not shown). That is why, the genetic structure was studied considering each multilocus genotype as two haplotypes. Haplotype reconstruction and inference of missing data were performed on the 1793 individuals using PHASE software v2.1 (Stephens *et al.* 2001) using the 17 microsatellite markers and the four polymorphisms of

candidate genes simultaneously. As suggested in Garrick *et al.* (2010), the model with recombination was used. The analysis was performed with 100 burn-in periods before 100 iterations, and a recombination rate between loci equal to 0.5. Then, pairs of haplotypes with the highest probability for each individual were selected. Focusing on the 17 microsatellite markers, this new data set was called phased multilocus genotypes. Among- and within-population haplotype variation was calculated using ARLEQUIN software v3.11 (Excoffier & Lischer 2010) by estimating unbiased haplotype diversity ( $H_D$ ), which accounts for small sample sizes (Nei 1987).

Two methods were used to detect the structure of genetic diversity based on the phased multilocus genotypes: (i) a discriminant analysis on principal component (Jombart *et al.* 2010) and (ii) a haplotype network analysis specifically adapted for the purpose of the study. TBB individuals, which had been already identified due to their high level of divergence compared to the other varieties, were not included in the analysis. In addition, a few very similar individuals showed a distinctly different pattern from the rest (7 individuals). They were identified as Florence-Aurore (FLA), an old variety also cultivated by the farmer who initiated the mixture. It was detected using data at 13 of the 17 microsatellite markers that were also used in a previous diversity study where FLA was genotyped (Roussel *et al.* 2004). These individuals were also discarded from the analysis.

The discriminant analysis was run using ADEGENET v1.2-6 (Jombart 2008), a package developed in R v3.1.2 (R Core Team 2014). It was applied to the results of a k-means clustering (Hartigan & Wong 1979). K-means is an unsupervised classification procedure that aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. K-means was performed on data obtained after a principal component analysis on centred data. Different numbers of initial seeds and of iterations were tested for a range of group numbers (from  $k = 2$  to 10) with 10 replications each time. The k-means algorithm was the most convergent one for 160 initial seeds and 100 000 iterations for each value of  $k$ . The  $k$  value was chosen using the Bayesian information criterion. Then, the discriminant analysis was carried out on the principal components, using the optimal number of groups detected previously. This clustering provided the probability of assignment of each individual to each of the  $k$  groups.

Haplotype network analysis, an adaptation of Rozenfeld's method (Rozenfeld *et al.* 2008), was performed using the number of differences among all pairs of unique haplotypes present in the data set (Thomas *et al.* 2012). These data were stored in a matrix ( $A$ ) where  $A_{ij}$

is the number of differences between haplotypes  $i$  and  $j$ . Undirected networks were plotted where each node corresponded to a distinct haplotype and edges linked two haplotypes  $i$  and  $j$  only if  $A_{ij} \leq th$ , with  $th$  the maximum number of differences between two haplotypes,  $th$  varying from 1 to 17. Networks were drawn with PAJEK software v1.26 (Batagelj & Mrvar 2002). Kamada–Kawai’s force-based algorithm (Kamada & Kawai 1989) was used to define the spatial distribution of the nodes. The threshold was set to 2 so as to detect haplotypes specific to each of the initial four varieties (GBTAN, GBTBB, GBTBP and GBTBSB) as independent genetic groups (TAN, TBB, TBP and TSB, respectively). Then, each haplotype unspecific to the four initial varieties but connected to one of the four genetic groups was assigned to this group. In this context, a genotype group is characterized by a network of frequent haplotypes connected to numerous rare haplotypes. A new genetic group was defined for every set composed of a minimum of five distinct haplotypes. Other haplotypes were defined as unassigned haplotypes. The analyses were implemented in R v3.1.2 (R Core Team 2014). The scripts are available on request.

The results of the two clustering methods (discriminant and haplotype network analyses) were compared and summarized in the form of pseudo-alleles corresponding to the different genetic groups, and called virtual multi-allelic marker. To connect genetic data and phenotypes, haplotype data were assembled into genotypes. To that end, we assigned each individual to a genotype group (GENOGP) resulting from the two pseudo-alleles. The GENOGP of each individual was defined based on its two pseudo-alleles. For instance, if the genetic group TBB was detected twice in the same individual, then its genotype group was set to TBBTBB. Only heterozygotes falling into two distinct genetic groups were defined as heterozygous for this marker. The virtual multi-allelic marker was used to monitor the evolution of the composition of each population in terms of group frequencies.

*Within- and among-population diversity.* Genetic diversity of the 1793 phased multilocus genotypes distributed among the 21 MDT samples and among the four initial varieties was studied at the level of the 17 microsatellite markers, at the virtual multi-allelic marker level defined in the previous section and at the level of the four polymorphisms in candidate genes (*FTA*, *VRN-1A<sub>prom</sub>*, *VRN-1A<sub>ex7</sub>* and *VRN-1D*). Unbiased Nei’s estimate of genetic diversity ( $H_E$ ) (Nei 1978), mean observed heterozygosity ( $H_O$ ), allele richness ( $R_S$ ) and deviation from Hardy–Weinberg genotypic proportions ( $F_{IS}$ ) were estimated with GENETIX software v4.05.2 (Belkhir *et al.* 2000). Haplotype richness ( $HR$ ) was computed as the number

of unique haplotypes in a population divided by twice the number of individuals in the population. Haplotype diversity ( $H_D$ ) was computed as the equivalent of  $H_E$  for haplotypes (Nei 1987).

Differentiation among populations was calculated using  $\theta$ , the  $F_{ST}$  estimator developed in Weir & Cockerham (1984) and implemented in Genetix software (Belkhir *et al.* 2000).

*Estimation of the genetic effective size.* The genetic effective size ( $N_E$ ) was estimated using the temporal variation in allele frequencies for each of the 17 microsatellite markers (Waples 1989).  $N_E$  was given by:

$$N_E = \frac{t_y - t_x}{2(\hat{F}_c - 1/S_x - 1/S_y)} \quad \text{eqn 1}$$

where  $S_x$  is the number of individuals sampled at generation  $t_x$  (respectively,  $S_y$  individuals at  $t_y$ ) and  $\hat{F}_c$  is the variance in allele frequency defined by:

$$\hat{F}_c = \frac{\sum_l K_l \hat{F}_{c,l}}{\sum_l K_l} \quad \text{eqn 2}$$

where  $K_l$  is the number of alleles at locus  $l$  and  $\hat{F}_{c,l}$ :

$$\hat{F}_{c,l} = \left( \frac{1}{K_l} \right) \sum_{i=1}^{K_l} \frac{(p_{x(i,l)} - p_{y(i,l)})^2}{(p_{x(i,l)} + p_{y(i,l)})/2 - p_{x(i,l)}p_{y(i,l)}} \quad \text{eqn 3}$$

where  $p_{x(i,l)}$  [respectively,  $p_{y(i,l)}$ ] represents the frequency of allele  $i$  at locus  $l$  in the sample of  $S_x$  individuals drawn at generation  $t_x$  (respectively,  $S_y$  individuals at  $t_y$ ).

*Statistical analyses of earliness.* As our experimental conditions were quite stressful for the plants, Fisher’s exact test was performed on the distribution of the number of viable plants and headed plants to detect whether certain genotype groups or populations were more specifically affected by growth conditions. These tests were performed using R (R Core Team 2014). SAS/STAT software, version 9.2 of the SAS System for Unix (Copyright © 2002–2008 SAS Institute Inc), was used for the other statistical analyses. Analyses of covariance (ANCOVA) were performed using the generalized linear model (GLM) procedure. The normal distribution of the residuals was tested using the UNIVARIATE procedure, and variances were estimated using the VARCOMP procedure with the restricted maximum-likelihood (REML) approach.

The basic initial model addressed the within-trial micro-environmental variations due to the experimental conditions. To account for potential competition among neighbouring plants, a neighbourhood covariate (NBH) was computed for each plant as the difference between the mean plant height of the eight neighbouring plants and the plant height of the considered individual plant.



The effects of the replicated block (*REP*) and of neighbourhood (*NBH*) were tested as follows:

$$Y_{kl} = \mu + \alpha NBH + REP_k + \varepsilon_{kl} \quad \text{eqn 4}$$

where  $Y_{kl}$  is the heading date value of plant  $l$  in  $REP_k$ ,  $NBH$  is a continuous variable,  $REP$  is the fixed block effect, and  $\varepsilon_{kl}$  is the random residual variable ( $\varepsilon \sim \mathcal{N}(0, \sigma_R^2)$ ).

The environmental variance ( $\hat{\sigma}_E^2$ ) of the trial was estimated as the residual variance of the following model:

$$Y_{ikl} = \mu + GENO_i + \alpha NBH + REP_k + \varepsilon_{ikl} \quad \text{eqn 5}$$

where the random  $GENO$  ( $i = 1..11$ ) effect corresponds to the 11 most frequent homozygous genotypes ( $GENO$ ) in the whole data set. Thus, the residual variance of this model corresponded to between-plant environmental variability.

*Spatial differentiation.* We then tested for differentiation among populations within the MDT snapshot using the following model:

$$Y_{ikl} = \mu + POP_i + \alpha NBH + REP_k + \varepsilon_{ikl} \quad \text{eqn 6}$$

where  $POP$  ( $i = 1..15$ ) is the random population effect corresponding to the 15 MDT populations sampled in 2007 and 2008.

In this model, the estimated variance of the  $POP$  effect gave the among-population genetic variance ( $\hat{\sigma}_{Gamong}^2$ ). As the residual variance of the model included both within-population genetic variance ( $\hat{\sigma}_{Gwithin}^2$ ) and among-plant environmental variation ( $\hat{\sigma}_E^2$ ),  $\hat{\sigma}_{Gwithin}^2$  was obtained as follows:

$$\hat{\sigma}_{Gwithin}^2 = \hat{\sigma}_{Residual}^2 - \hat{\sigma}_E^2 \quad \text{eqn 7}$$

with  $\hat{\sigma}_E^2$ , the environmental variance estimated in model (5).

To compare differentiation at a quantitative trait level to differentiation at the neutral marker level ( $F_{ST}$ ),  $Q_{ST}$  (Wright 1969; Spitze 1993) was estimated for the MDT snapshot subset:

$$Q_{ST} = \frac{(\hat{\sigma}_{Gamong}^2(1 + F_{IS}))}{(\hat{\sigma}_{Gamong}^2(1 + F_{IS}) + 2\hat{\sigma}_{Gwithin}^2)} \quad \text{eqn 8}$$

As wheat is mainly a selfing species, we assumed that  $F_{IS} \simeq 1$ , and we obtained:

$$Q_{ST} = \frac{(\hat{\sigma}_{Gamong}^2)}{(\hat{\sigma}_{Gamong}^2 + \hat{\sigma}_{Gwithin}^2)} \quad \text{eqn 9}$$

A multiple regression model was used to assess whether the average heading date of the different populations could be explained by environmental characteristics of the cultivation sites such as latitude ( $LAT$ ), longitude ( $LONG$ ), altitude ( $ALT$ ), demographic

size ( $SIZE$ ) and the number of generations during which the mixture had been grown on a given farm after moving from A's ( $NBGENOUT$ ). The multiple regression was performed using a stepwise model with the REG procedure and the FORWARD method.

The role of genetic structure in earliness differentiation was investigated at the population level. A regression analysis was carried out to test whether the frequency of the genotype groups composing the samples explained the average heading date.

The effect of genetic structure was also investigated at the individual level to test whether phenotypic differences were due to the genotype groups:

$$Y_{mkl} = \mu + GENOGP_m + \alpha NBH + REP_k + \varepsilon_{mkl} \quad \text{eqn 10}$$

with  $GENOGP$  a random effect ( $m$  in 1..7), characterizing each of the main genotype groups detected in section *Detection of the genetic structure* (discriminant analysis and haplotype network analysis) to limit the number of unbalanced classes.

Then, the model (6) was run on data subsets composed of each main genotype group to estimate among-population within-group genetic variance ( $\hat{\sigma}_{Gamong}^2$ ) for each of them and to compare it with their genetic diversity ( $H_E$ ).

*Association study between heading date and candidate genes for earliness.* The effects of candidate gene polymorphisms ( $MK_j$ ) on the heading date were assessed at three levels, considering :

- the population effect:

$$Y_{ijkl} = \mu + POP_i + MK_j + \alpha NBH + REP_k + \varepsilon_{ijkl} \quad \text{eqn 11}$$

- the genetic background ( $GENOGP$  effect):

$$Y_{ijkl} = \mu + GENOGP_i + MK_j + \alpha NBH + REP_k + \varepsilon_{ijkl} \quad \text{eqn 12}$$

- the population effect within a given genotype group: the model (11) was run for each of the main genotype group.

Each polymorphism ( $MK$ :  $VRN-1A_{prom}$ ,  $VRN-1D$ ,  $VRN-1A_{ex7}$  and  $FTA$ ) was tested both separately and altogether.

## Results

### *Multilevel genetic structure of the MDT populations*

Virtual MDT did not show significantly different diversity at the neutral marker level (expected heterozygosity,  $H_E = 0.49$ , and haplotype diversity,  $H_D = 0.88$ ) as



compared to A03 ( $H_E = 0.51$  and  $H_D = 0.94$ , Table S1 in Supporting Information). The number of shared haplotypes between virtual MDT and A03 was quite low (only seven distinct haplotypes) with 82.9% of the haplotypes initially present in virtual MDT not detected in A03. Conversely, A03 population was composed of 88.7% of haplotypes not present in virtual MDT, among which 61.3% were recombinant haplotypes among haplotypes present in virtual MDT and 27.4% were new haplotypes with one to three new alleles per individuals undetected in virtual MDT (Fig. 2a). Consistently, a significant pairwise  $F_{ST}$  was estimated between virtual MDT and A03 ( $F_{ST} = 0.05$ ,  $P$ -value $<0.001$ ). Genetic diversity at the MDT snapshot level ( $H_E = 0.51$ ) was similar to initial diversity in A03 ( $H_E = 0.51$ ) (Table S1 in Supporting Information). The MDT snapshot showed slightly higher haplotype diversity than A03 ( $H_D = 0.98$ ) and higher allelic richness ( $R_S = 9.35$  vs.  $R_S = 3.65$  for A03). Only 14 distinct haplotypes among the 440 detected in the MDT snapshot were shared with virtual MDT. Within the 97% of haplotypes of the MDT snapshot not present in virtual MDT, 7% were shared with A03, 34% were recombinants of virtual MDT, 15% were recombinants of A03, and 41% were haplotypes with new alleles present neither in virtual MDT nor in A03 (Fig. 2b). A significant pairwise  $F_{ST}$  was estimated between A03 and the MDT snapshot ( $F_{ST} = 0.09$ ,  $P$ -value $<0.001$ ). The genetic parameters of the MDT populations revealed two groups of populations. Five populations (C08, I08-1, I08-2, E08 and K08) displayed lower diversity than A03 and the MDT snapshot for  $H_E$  and  $H_D$ , while the other had the same level of diversity as farmer A's populations (Table S1 in Supporting information).

*Genetic structure of the MDT populations.* A clustering analysis on the 1793 pooled phased multilocus genotypes was performed at the haplotype level working with 517 unique haplotypes.

The  $k$ -means method revealed that the minimal BIC value was obtained for  $k = 6$  groups, with a strong and stable elbow compared to the other  $k$  values (figure not shown). Two genetic groups were easily assigned to TAN and TBP on the basis of the reference varieties. The sample of the TSB reference variety for TSB was subdivided into two groups (further named TSB2 and TSB3). At this stage, two other groups were still not assigned.

The haplotype network also allowed us to detect TAN and TBP (Fig. 3), but for TSB only one main group and a closely related minor one (named TSB5) were detected. A new subgroup was detected within the TSB3 group. It was only observed in the E08 population. Interviews revealed that in 2006, this population was exposed to a mixture with another landrace called

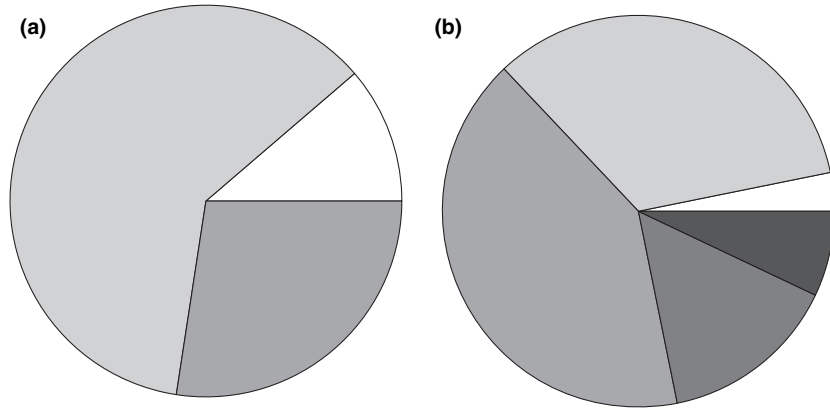
*Touselle des Hautes-Alpes*. Therefore, this TSB group was assumed to be derived from this landrace and was called THA even though no sample was available to check whether we actually detected that particular variety. The remaining TSB3 individuals were maintained as the TSB3 group. Combining both discriminant analysis and haplotype network approaches led to the conclusion that the two unknown groups detected with discriminant analysis most likely fell into the TSB variety. They will further be named TSB1 and TSB4.

Haplotypes belonging to the same genetic group (of the same colour) appeared connected and close to each other, while most of the unassigned haplotypes were on the outskirts of the network, consistent with the purpose of the method (Fig. 3). The haplotype network method allowed us to detect each haplotype that did not fall into one of the genetic groups. They were labelled as unassigned haplotypes, acknowledging that disentangling recombinants, migrants or experimental artefacts is a complex task. Further investigations would be needed to determine the nature of these haplotypes. In this study, we focused on the haplotypes belonging to the 10 identified genetic groups: TAN, TBB, TBP, TSB1, TSB2, TSB3, TSB4, TSB5, THA and FLA. They represented 94.0% of the 2234 haplotypes sampled within the MDT snapshot, corresponding to 96% of the 1793 individuals. Unassigned haplotypes are not considered in this study; therefore, the role of farmers' practices on the migration process among populations within a same farm will not be analysed.

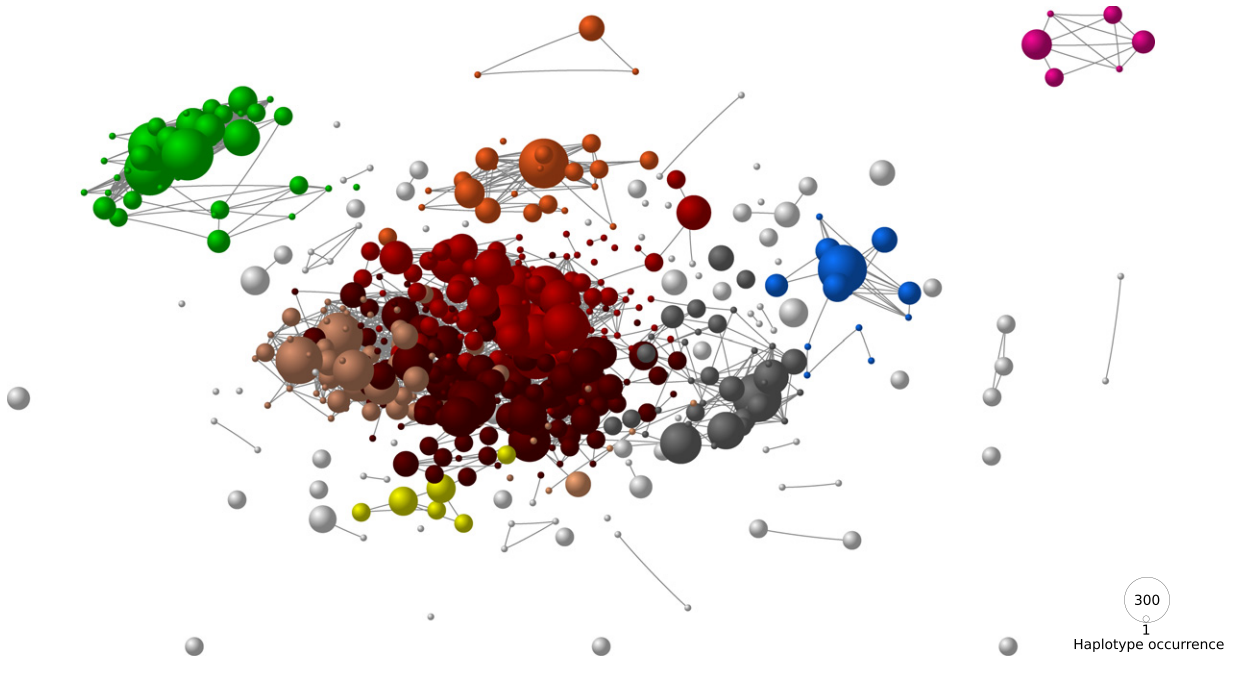
*Diversity of the initial four landraces.* We analysed the genetic diversity of the initial four landraces (Table 2) in terms of the eight genetic groups (TAN, TBB, TBP, TSB1, TSB2, TSB3, TSB4 and TSB5). Each landrace belonged to one genetic group except TSB which was divided into two groups, TSB2 and TSB3. TAN, TBB and TBP exhibited lower values of genetic diversity than TSB whatever the index ( $H_E$  or  $R_S$ ) and the sample (virtual MDT, HE03fld or the MDT snapshot, Table 2).

#### *Evolution of genetic composition across the MDT snapshot*

At the MDT snapshot scale, the proportions of the different groups were more balanced than in the reference population (A03). On average, TBB decreased in frequency while the TSB groups became more frequent, with a particular increase in TSB1. However, group composition varied drastically from one population to the other (Fig. 1). Sixteen of the 21 populations were composed of more than 50% of TSB groups. TAN was maintained at a low frequency in most of the populations except in C08, E08, F07, I08-1 and K08, where it



**Fig. 2** Distribution of the unique haplotypes observed in : (a) the initial reference population of *Mélange de Touselles* (A03) (62 haplotypes), (b) *Mélange de Touselles* (MDT) snapshot (440 haplotypes). Shared haplotypes with Virtual MDT (corresponding to the four reference varieties pooled together: GBTAN, GBTBB, GBTBP and GBTBB) are in white. Recombinant haplotypes of virtual MDT are in light grey. Haplotypes with new allele(s) are in medium grey. Shared haplotypes with A03 are in dark grey. Recombinant haplotypes of A03 are in black.



**Fig. 3** Global haplotype network composed of 517 distinct haplotypes (nodes) detected after pooling all together the different *Mélange De Touselles* (MDT) samples. Two haplotypes are connected if they show two or less differences. Spatial distribution of the nodes accounts for the total number of differences between each pair of haplotypes. Each colour corresponds to a different genetic group. *Touselle Blanche de Provence* (TSP) is in grey, TSB is divided into five subgroups (TSB1: orange, TSB2: red, TSB3: dark brown, TSB4: light brown and TSB5: salmon pink). *Touselle Anone* (TAN) is in blue. *Touselle Blanche Barbue* (TBB) is in green. *Florence-Aurore* (FLA) is in pink, *Touselles des Hautes-Alpes* (THA) in yellow and unassigned in white. Node size is proportional to the number of haplotypes found in the pooled data set (1793 phased multilocus genotypes, i.e. 3586 haplotypes).

was not detected. TBP was present at a low frequency in most of the populations, except in the B08 population where it was much higher (30%), and in C08, E08, I08-1, I08-2, K08 where it was not detected.

Pure temporal evolution of MDT was only captured by the time series from farmer A's population. The results indicated a nonsignificant temporal differentiation of MDT after 5 generations on A's farm :

**Table 2** Summary statistics to characterize: (i) the four reference varieties pooled together: GBTAN, GBTBB, GBTBP and GBTBSB (virtual MDT); (ii) the initial reference population of *Mélange de Touselles* (A03); (iii) the *Mélange de Touselles* snapshot (MDT snapshot). Results are detailed for the four landraces considering each haplotype as one homozygote individual. N: number of haplotypes,  $H_E$ : unbiased expected heterozygosity,  $R_S$ : average number of alleles per locus

Variety	Haplotype Group	Virtual MDT			A03			MDT snapshot		
		N	$H_E$	$R_S$	N	$H_E$	$R_S$	N	$H_E$	$R_S$
TAN	=	64	0.010	1.060	4	0.000	1.000	89	0.040	1.880
TBB	=	64	0.030	1.240	74	0.050	1.350	417	0.050	2.350
TBP	=	63	0.030	1.290	8	0.080	1.240	153	0.060	2.180
TSB		64	0.260	2.180	74	0.340	2.820	1543	0.340	5.470
—	TSB1	0	—	—	18	0.170	1.590	634	0.160	3.350
—	TSB2	6	0.050	1.120	4	0.060	1.120	158	0.040	2.470
—	TSB3	58	0.180	1.710	28	0.210	1.760	457	0.220	3.940
—	TSB4	0	—	—	22	0.100	1.470	286	0.110	2.590
—	TSB5	0	—	—	2	0.040	1.060	8	0.120	1.350

$F_{ST}(\text{multilocus}) = 0.014$  and  $F_{ST}(\text{VMK}) = 0.011$  (lower bound of  $F_{ST}$  estimations were below 0).

The other sampled populations corresponded to independent evolutions of the initial A03 mixture. Differentiation estimated at the MDT snapshot level [ $F_{ST}(\text{multilocus}) = 0.123$ ,  $F_{ST}(\text{VMK}) = 0.158$ ] indicated a significant differentiation among the set of populations harvested in 2007–2008 (lower bound of  $F_{ST}$  estimation always higher than 0).

#### Genetic effective population size

The effective size estimated in farmer A's farm fluctuated from 10 to 30 individuals when it was possible to make  $N_E$  estimation (Table S2 in Supporting information). Almost all populations displayed an  $N_E$  value of the same order of magnitude as farmer A's population, that is ranging between 5 and 32, whatever the number of generations. M08, D08, H08, N08 and D06 samples showed a higher effective size ranging between 62 and 153. As confidence intervals did not overlap these four populations,  $N_{ES}$  were considered as significantly higher than the rest of the samples except for A05, A08 and L08 (Table S2 in Supporting information).  $N_E$  was not estimated for A06, N07 and G08 because  $\hat{F}_C$  variation was too small compared to sample size, leading to values tending towards infinity. In this particular case, allele frequencies were stable after the circulation event, unlike in the other populations.

#### Evolution of earliness

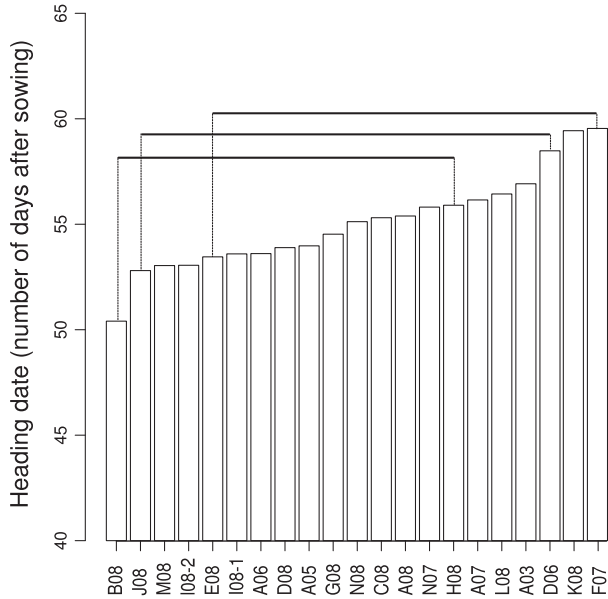
*Phenotypic evolution.* To connect genetic data and phenotypes, genotype groups instead of genetic groups will be considered in the following paragraphs. As our

experimental conditions were quite stressful for the plants, 25% did not reach heading stage. The two Fisher's tests on the distribution of the number of viable plants and headed plants were highly significant ( $P_{value} < 0.0001$ ) for the genotype group (GENOGP) effect as well as for the population (POP) effect. The TBBTBB group was the most severely affected by these particular conditions (28.5% of the TBBTBB plants did not reach heading stage, compared with 25% overall).

The simple model (eqn 4), with a REP effect and the neighbourhood covariate (NBH), explained 1.27% of total variation in the heading date. REP effect was not significant, whereas NBH covariate was highly significant ( $P_{value} < 0.0001$ ), with a positive coefficient ( $\hat{\alpha} = 0.79$ ) indicating that competition due to plant height differences among neighbouring plants significantly delayed heading.

*Mélange de Touselles* snapshot differentiation was studied with model (eqn 6), using all the populations sampled in 2008, plus F07 in 2007. The POP effect was highly significant ( $R^2 = 12.7\%$ ,  $P_{value} < 0.0001$ ), indicating among-population divergence for earliness. B08, J08, M08 and I08-1 populations were significantly earlier than K08 and F07 (with an average 6.4–9.1 days' delay, Fig. 4). Earliness in the other populations ranged between these extremes.  $\hat{\sigma}_{\text{Gamong}}^2$  and  $\hat{\sigma}_{\text{Gwithin}}^2$  (eqn 7) were estimated to compute  $Q_{ST}$  (eqn 9). The  $Q_{ST}$  value (0.26) was much higher than the multilocus  $F_{ST}$  value (0.11). This result indicates that differentiation was faster and higher for earliness than for neutral markers.

The best multiple regression model explained 7.0% of variability. The effect of demographic size was not significant, while latitude, longitude, altitude and the number of generations outside farmer A's farm were significant. Latitude was the most significant parameter

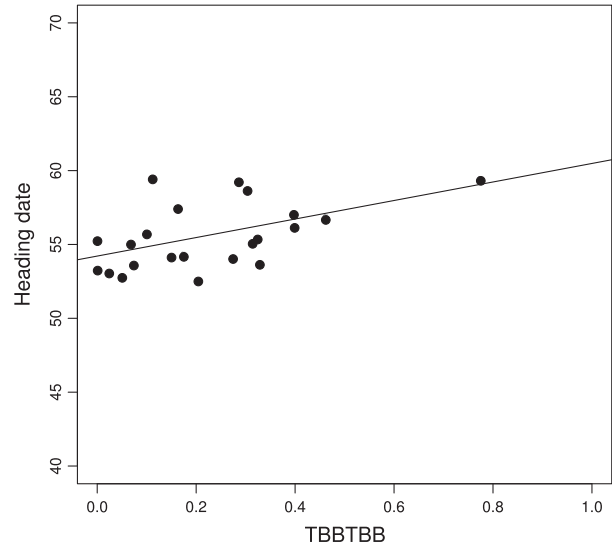


**Fig. 4** Least squares means for heading date calculated for each population. The three lines above the bars indicate populations sharing the same statistical group.

(partial  $R^2 = 2.4\%$ ), with a positive value for the estimated coefficient. This indicated that populations located in the south became slightly earlier than populations grown in the north. Therefore, a large part of among-population earliness differentiation was due to other factors.

*Genetic bases of heading date differentiation.* In the regression of the mean heading date of the populations on the frequency of each genotype group within populations, only TBBTBB had a significant positive effect (Fig. 5). But in general, the mean heading dates of the populations were only weakly influenced by the frequencies of the genotype groups.

The effect of genetic structure was tested on individual heading dates using the genotype groups data. Only the seven most frequent genotype groups (TANTAN, TBBTBB, TBPTBP, TSB1TSB1, TSB2TSB2, TSB3TSB3 and TSB4TSB4) were considered (*GENOGP*). Our selection discarded heterozygotes that belonged to different genetic groups, but did not discard heterozygotes that belonged to a same genetic group. Including *GENOGP* instead of *POP* in the analysis of covariance (eqn 10 vs. eqn 6) dramatically increased the explanation in Table 3. The effect of *GENOGP* was highly significant ( $P_{value} < 0.0001$ ), with substantial differences among weighted means (LSmeans) for genotype groups (Table 3 and Fig. 6). TSB2TSB2 was the earliest genotype group ( $HD = 48.5$ ), while TBBTBB was the latest ( $HD = 60.0$ ), and the others (TSBTSBs, TBPTBP and TANTAN) exhibited intermediate values (Fig. 6).



**Fig. 5** Correlation between least squares means values of heading dates per population and the frequency of the genotype group TBBTBB in each population ( $R^2 = 0.24$  and  $P_{value} = 0.012$ ).

Finally, we studied among-population variation for each genotype group (eqn 6, Table 4). TSB2TSB2 showed the highest among-population genetic variance ( $\hat{\sigma}_{\text{Gamong}}^2 = 8.85$ ) and TBPTBP the lowest ( $\hat{\sigma}_{\text{Gamong}}^2 = 0.00$ ), with a nonsignificant *POP* effect.

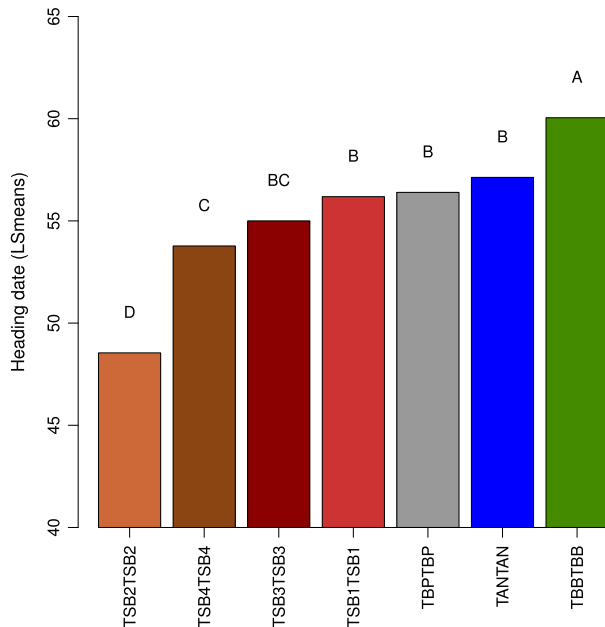
*Differentiation at the level of earliness candidate genes.* Genetic diversity in earliness candidate genes was altogether lower than at the neutral marker level with similar trends across populations (Table S1 in Supporting information).

Differentiation in the candidate genes associated with earliness ( $F_{STQ} = 0.239$ ) was similar to differentiation for the quantitative trait ( $Q_{ST} = 0.261$ ) and much larger than neutral allelic differentiation ( $F_{ST}(\text{multilocus}) = 0.111$ ) and genotype group differentiation ( $F_{ST}(\text{virtual}) = 0.158$ ). Phenotypic differentiation might be partly underlain by differentiation at *FTA*, *VRN-1A* and *VRN-1D*.

*Association between heading date and candidate genes.* The effects of each gene polymorphism were tested one by one as well as together with the *POP* effect in eqn 11. These models accounted for 16.0–36.5% of variability, with a strong significant *POP* effect. *FTA* and *VRN-1A<sub>ex7</sub>* were not significant, whereas *VRN-1D* and *VRN-1A<sub>prom</sub>* were highly significant when tested one by one ( $P_{value} < 0.0001$ , data not shown). These two genes were generally associated with the heading date whatever the population. Only *VRN-1A<sub>prom</sub>* remained significant when the four genes were tested all together. The

**Table 3** ANCOVA table for heading date considering the effect of the 15 MDT populations sampled in 2007 and 2008 (*POP*) and the effect of the most represented genotype groups (*GENOGP*). Experimental artefacts were controlled by the block effect (*REP*) and the neighbourhood covariate (*NBH*). Significance of the tests are given by : \*\*\*:  $P_{value} < 0.0001$ , —: not computed in the current model

Model	<i>NBH</i>		<i>REP</i>		<i>POP</i>		<i>GENOGP</i>		$R^2$
	d.f.	<i>F</i> -value	d.f.	<i>F</i> -value	d.f.	<i>F</i> -value	d.f.	<i>F</i> -value	
(6)	1	15.05***	1	0.83	21	7.77***	—	—	0.12
(10)	1	54.63***	1	0.79	—	—	6	77.08***	0.30



**Fig. 6** Weighted means for heading dates estimated for each genotype group. Significance of the tests was provided after Bonferroni's multiple correction.

model (eqn 12) allowed to assess the association between earliness genes and heading date while accounting for the genetic structure of the genotype groups. The models explained between 28.3 and 30.7% of variability, and *GENOGP* was always highly significant ( $P_{value} < 0.0001$ ) (Table 5). The effects of *VRN-1A<sub>ex7</sub>* and *VRN-1D* were not significant, whereas *FTA* and *VRN-1A<sub>prom</sub>* were significant ( $P_{value} < 0.05$  and  $P_{value} < 0.0001$ , respectively). The *FTA* marker was not in association in the previous model due to significant interactions between genotype groups and *FTA* alleles (data not shown). TSB1TSB1, TSB3TSB3 and TSB4TSB4 preferentially carried allele 1 of *FTA*, whereas TANTAN, TBBTBB, TBPTBP and TSB2TSB2 carried allele 2. Nevertheless, the same average heading date value was observed for these two allele-specific groups. The model (eqn 11) was also tested in each genotype group. *FTA* was significant in TSB1TSB1 and TSB3TSB3, and *VRN-*

*1A<sub>prom</sub>* was significant in TSB2TSB2 (Table 6). Individuals from TSB1TSB1 and TSB3TSB3 carrying *FTA* allele 2 were earlier than individuals from the same group carrying allele 1 (3 days for TSB1TSB1, and more than 5 days for TSB3TSB3, Fig. 7). In TSB2TSB2, individuals carrying allele 1 of *VRN-1A<sub>prom</sub>* headed 5 days earlier than those carrying allele 2 (Fig. 7).

## Discussion

Our snapshot study of MDT, a farmers' mixture of wheat landraces, evidenced significant differentiation patterns at the levels of neutral markers, of candidate genes involved in earliness and of earliness, a quantitative adaptive trait. Population differentiation was influenced by combined demographic and selection processes. We did not analyse the role of migration processes but rather controlled for the effect of migration on differentiation by detecting the new genotypes potentially introduced by migration.

### Initial structure of the MDT mixture

Mélange de Touselles displayed a tricky population structure, with populations composed of different varieties of two species that were in turn composed of several genetic groups with some overlaps among varieties within the same species. Among the four components of MDT, TBB, the *Triticum turgidum* landrace and TAN showed low genetic diversity compared to TSB and TBP, with TBP embedded within TSB. Moreover, the genetic structure of TBP and TSB was similar to that of Rouge de Bordeaux, another on-farm-managed population variety of bread wheat (Thomas *et al.* 2012). These landraces were composed of a few major haplotypes connected to much less frequent haplotypes in the periphery. Relatedness among genotype groups also indicated that they shared part of their genetic background although they had specific alleles that allowed us to distinguish among them using control individuals.

The impact of the founder effect in the initial farm was analysed by comparing the composition of the on-

**Table 4** ANCOVA table for heading date obtained with model (6) run for each genotype group and considering *POP* effects. \*:  $P_{value} < 0.05$ , \*\*:  $P_{value} < 0.01$  and \*\*\*:  $P_{value} < 0.0001$

Genotype group	<i>NBH</i>		<i>REP</i>		<i>POP</i>		Residual		$R^2$
	d.f.	<i>F</i> -value	d.f.	<i>F</i> -value	d.f.	<i>F</i> -value	$\hat{\sigma}_{\text{Gamong}}^2$	$\hat{\sigma}^2$	
TANTAN	1	0.89	1	0.00	15	1.42	4.70	21.52	0.45
TBBTBB	1	18.57***	1	0.31	19	4.29***	8.31	18.75	0.32
TBPTBP	1	5.86*	1	0.18	16	1.46	0.00	22.74	0.30
TSB1TSB1	1	6.62*	1	0.67	21	2.92***	2.89	25.41	0.19
TSB2TSB2	1	6.93**	1	0.88	19	2.78**	8.85	25.18	0.38
TSB3TSB3	1	3.12	1	0.08	21	1.58*	1.22	20.81	0.15
TSB4TSB4	1	3.20	1	0.34	19	2.65**	7.56	23.07	0.32

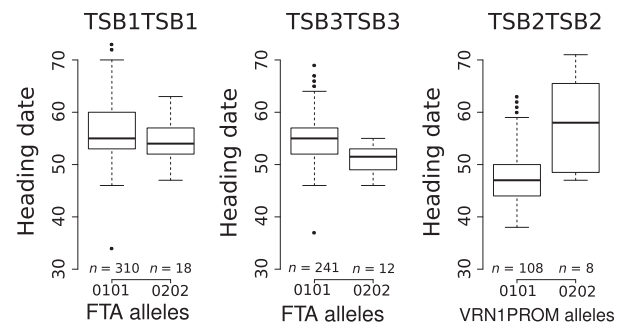
**Table 5** ANCOVA table with the *F* statistics, the significance of the tests and the variability in the response explained by the explanatory variables ( $R^2$ ) for heading date obtained with model (11) and (12). The following factors were tested: the effect of the 15 MDT populations sampled in 2007 and 2008 (*POP*), the effect of the most represented genotype groups (*GENOGP*) and the effects of the different candidate genes (*FTA*, *VRN-1A<sub>ex7</sub>*, *VRN-1D*, *VRN-1A<sub>prom</sub>*). Experimental artefacts were controlled by the block effect (*REP*) and the neighbourhood covariate (*NBH*). Significance of the tests is given by \*:  $P_{value} < 0.05$  and \*\*\*:  $P_{value} < 0.0001$ . —: not computed in the current model

Model	<i>NBH</i>	<i>REP</i>	<i>POP</i>	<i>GENOGP</i>	<i>FTA</i>	<i>VRN-1A<sub>ex7</sub></i>	<i>VRN-1D</i>	<i>VRN-1A<sub>prom</sub></i>	$R^2$
(11)	66.93***	0.55	7.47***	—	1.06	2.03	0.34	26.16***	0.37
(12)	54.47***	0.24	—	9.83***	3.66*	1.95	0.03	11.79***	0.31

**Table 6** ANCOVA table with the *F* statistics and the significance of the tests and the variability in the response explained by the explanatory variables ( $R^2$ ) for heading date obtained with model (11) for TSB1TSB1, TSB2TSB2, TSB2TSB2, the genotype groups with significant effect for at least one of the earliness genes. The following factors were tested: the effect of the 15 MDT populations sampled in 2007 and 2008 (*POP*) and the effects of the different candidate genes (*FTA*, *VRN-1A<sub>prom</sub>*). Experimental artefacts were controlled by the block effect (*REP*) and the neighbourhood covariate (*NBH*). Significance of the tests is given by \*:  $P_{value} < 0.05$ , \*\*:  $P_{value} < 0.01$  and \*\*\*:  $P_{value} < 0.0001$ , —: not computed in the current model

Genotype group	<i>NBH</i>	<i>REP</i>	<i>POP</i>	<i>FTA</i>	<i>VRN-1A<sub>prom</sub></i>	$R^2$
TSB1TSB1	6.42*	0.8	3.01***	5.91*	—	0.21
TSB3TSB3	2.23	0.13	1.54	7.95**	—	0.17
TSB2TSB2	9.55**	2.12	2.37**	—	4.8*	0.49

farm reference sample A03 to the virtual MDT. The two populations greatly differed, with only 11% of shared haplotypes. This was probably due to differences in the samples provided by the GenBank to farmer A in 1997 and to us in 2008. During the multiplication phase of *ex situ* wheat accessions, a limited number of seeds (60) are sown and a strong morphological selection is



**Fig. 7** Association between alleles and heading date for three genotype groups (TSB1TSB1, TSB3TSB3 and TSB2TSB2) and two genes (*FTA* and *VRN-1A<sub>prom</sub>*). Assessed genotypes come from the different *Mélange De Touselles* samples. Class sizes are mentioned under each boxplot.

applied on plants to conserve only one representative spike-type per accession. Moreover, only a few plants that correspond to the ‘type’ are self-pollinated to produce the regenerated sample, leading to strong genetic drift effects. Genetic drift effects have already been observed when within-landrace genetic diversity conserved *ex situ* for different time periods was compared to the diversity conserved on-farm [in barley (Parzies *et al.* 2000), bean (Gómez *et al.* 2005), pea (Leino *et al.*



2013)]. In addition, when people ask for accessions, the GenBank provides them with 40–50 seeds per accession.

Therefore, a moderate to strong bottleneck effect might occur at the initial level of within-accession genetic diversity, corresponding to the founder effect when the number of individuals involved in colonization is small (Slatkin 1977; Wade & McCauley 1988). Farmer A received samples of 40–50 seeds from the GenBank in 1997 for each of the four landraces. In this study, virtual MDT was obtained with seeds provided by the same GenBank after a last multiplication in 2004. Thus, each variety was multiplied two or three times between 1997 and 2004. Whereas the precise composition of the initial mixture was not available, interviews with farmers and the GenBank curator provided some pieces of evidence that seed management practices such as the quantity of seed used for multiplication or circulation played a crucial role on MDT evolution, by increasing the influence of genetic drift.

#### *Evolution of the MDT mixture under on-farm management*

After this first phase of mixture creation, samples of farmer A's MDT were distributed to different farmers, and MDT populations evolved as independent entities in each particular environment. The level of genetic diversity was the same at the global level ( $H_E = 0.51$ ) as within the initial population, while haplotype diversity and allele richness were much higher at the global level. Diversity in earliness genes followed the same patterns (Table S1 in Supporting information). These results illustrate that submitting several subpopulations to contrasted environmental conditions allowed a good maintenance of initial allelic diversity at the global level of the different versions of MDT. Such a phenomenon was empirically observed for wild inbred populations of *Leavenworthia* (Liu *et al.* 1998) and bread wheat populations maintained in a dynamic management design (Goldringer *et al.* 2006). In the absence of turnover (extinction/recolonization), the increase in genetic differentiation associated with the low pollen migration rates in highly selfing species is the main driver of the stabilization of genetic diversity at the global level (Ingvarsson 2002).

A particularly high differentiation was observed for neutral markers (Pannell & Charlesworth 1999) and could partly explain the observed pattern of differentiation among populations (0.11) for multilocus  $F_{ST}$  and (0.15) for genotype group  $F_{ST}$ . The impact of on-farm management practices will then greatly depend on whether farmers grow their populations in small plots (collection) or in fields, that is on demographic size. Genetic drift has been reported for wheat and maize

when farmers grew populations in small plots (Zhang *et al.* 2006; van Heerwaarden *et al.* 2010). For one farmer (I) who continuously grew MDT in very small plots (around 80 plants in I08-1), founder effect and genetic drift were combined, leading to one of the lowest genetic diversity ( $H_E = 0.35$ ) and effective size ( $N_E = 10$ ) values. These hypotheses could be rigorously tested using the theoretical framework developed by van Heerwaarden *et al.* (2010); Artoisenet & Minsart (2014). However, that would require adapting the model to integrate founder effect to better account for farmers' practices and social organization.

Differentiation within the MDT snapshot was assessed at three levels: (i) neutral markers, (ii) polymorphisms in genes associated with earliness and (iii) earliness at the phenotypic level. We focused on earliness for its important role in adaptation to environmental conditions, in particular to the interaction between sowing date (farming practice) and climate. Populations need to synchronize their reproductive cycle to climate conditions (Rhoné *et al.* 2008), which are different from one farm to another, and also depending on farming practices. Differentiation within the MDT snapshot was higher at the candidate gene level ( $F_{STQ} = 0.24$ ) and for earliness ( $Q_{ST} = 0.26$ ) than at the multilocus level ( $F_{ST} = 0.11$ ). This particular differentiation pattern,  $F_{ST} < F_{STQ} \leq Q_{ST}$ , was interpreted as the result of divergent selection (Merila & Crnokrak 2001; McKay & Latta 2002), given that there was no or limited gene flow among populations (Le Corre & Kremer 2012). This interpretation relying on theoretical study is consistent with the history of MDT. Different subpopulations of MDT were disseminated in different socio-climatic environments through seed circulation, leading to quite strong divergent selective pressures on the mixture. In addition, farmers' practices, such as plot isolation or off-type cleaning, aimed to limiting gene flow among crop populations. Local exceptions were observed but did not seem to affect the global differentiation pattern. A similar pattern was observed for maize populations grown in Mexico (Pressoir & Bertaud 2003b), although the underlying social and evolutionary processes were different mainly due to the different mating system. Moreover, in our MDT snapshot, the stronger differentiation of the candidate genes showed that 10 of the 15 populations faced different climatic and farming conditions inducing divergent selection in gene regions involved in adaptive traits such as earliness.

The evolution of the crop populations is controlled by the interaction between farmer-mediated and natural selection. In the case of MDT management by a farmers' network, direct farmer-led selection was rare. However, particular practices have been reported by some



farmers and could reinforce the effect of natural selection, such as late sowing or harvesting before the seeds of all genotypes were mature. Unfortunately, these singular practices were only qualitatively reported and specific to one particular year, making it difficult to include in a statistical model. It would be interesting to study whether other fitness-related traits were also differentiated among populations. Including concrete measurements of farming management practices in the experimental design is critical to understand crop evolution processes. After a longer period of MDT evolution, extinction and new colonization can be expected in some farms. Once equilibrium state has been achieved, theoretical results from the metapopulation model could be helpful to understand the evolutionary mechanisms that shape crop metapopulations (van Herwaarden *et al.* 2010) and by extension crop metapopulations.

### Genetic mechanisms involved in earliness differentiation

In the context of a self-pollinated species, selection could significantly affect gene diversity and also neutral diversity, in particular around the genes submitted to selection, through hitch-hiking effects and/or background selection (Ingvarsson 2002). The strong variation in the composition of the mixture was positively correlated with altitude (Fig. S2 in Supporting information). The specific environmental conditions occurring high up in the mountains might have affected the viability of some particular landraces. *Touselles* landraces are not expected to be adapted to mountain conditions as they were historically grown in southeastern France.

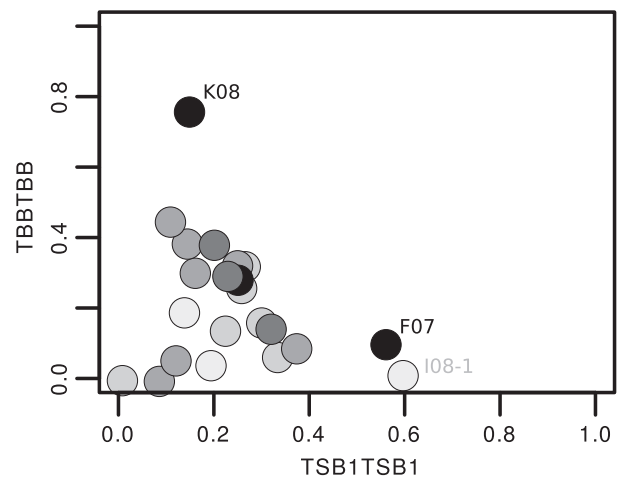
Therefore, we noticed that TBB, TAN and TBP, the three landraces that were the least diversified (Table 2), consistently decreased in frequency at high altitudes, while TSB increased. TSB was the only landrace component that showed a level of diversity ( $H_E = 0.26$ ) similar to the rare data available for other wheat landraces from Oman, Turkey and Mexico, that ranged between 0.15 and 0.55 (Dreisigacker *et al.* 2005; Zhang *et al.* 2006), while TAN, TBB and TBP had a lower level of diversity. Local adaptation of the mixture could result from two different processes shaping the standing variability available within the mixture: (i) selection among components of the mixture and (ii) selection within these components. These two processes are discussed below.

Strong differentiation among genetic groups was found for earliness. Thus, we assumed that changes in mixture composition could affect the average earliness value (Fig. 5 and Table 3). The positive correlation between TBBTBB frequency and the heading dates of populations suggested that at least part of phenotypic

divergence was due to variation in the proportion of TBBTBB. In addition to the evolution of the mixture composition, two landraces (TBB and TSB) showed significant among-population divergence for earliness, indicating within-landrace genetic evolution.

All landraces had a very low level of genetic diversity for earliness gene polymorphisms ( $H_e$  between 0 and 0.03, Table S1 in Supporting information). Allele fixation at earliness genes in these landraces induced statistical confusion between the effects of the genotype group and the gene. This could explain the lack of association between earliness and polymorphism at two of the four earliness genes in these populations, whereas a strong association was detected in a core collection of bread wheat (Rousset *et al.* 2011). Association between *FTA* and *VRN-1A<sub>prom</sub>* polymorphisms and earliness was detected within the TSB groups present in the MDT snapshot (Fig. 7, Table 4). Therefore, genetic evolution within landraces in the different mixtures seemed to play a significant role in the differentiation of populations for heading date. This is consistent with the fact that TSB was the most frequent in many samples and was maintained at a very high level in the MDT snapshot, indicating higher adaptability. Moreover, such a body of evidence indicates that TSB still kept the ability to adapt to contrasted environments. Specific additional experiments would be needed to confirm this hypothesis.

Lastly, we found that different genetic trajectories led to the same heading date at the population level, as illustrated in Fig. 8. For instance, two populations were late heading with a high frequency of TBBTBB (K08), but also with a very low frequency of TBBTBB (F08).



**Fig. 8** Relationship between the frequencies of two genotype groups TSB1TSB1 and TBBTBB and the mean heading date for each *Mélange De Touselles* samples (points). Heading date values are represented by the different colours: dark colours represent late heading dates and light colours represent early heading dates.

Paradoxically, the same frequency of TBBTBB and TSB1TSB1 in another population was associated with an early heading date (I08-1, light colour). These findings provide an example of the strong genetic ‘plasticity’ of mixtures and their ability to adapt to different environments. Despite the challenges posed by the on-farm evolutionary experiment where populations are submitted to real farming conditions, we were able to highlight how standing genetic variation plays a key role in adaptive processes.

## Conclusion

This study aimed to depict the main genetic characteristics of a recently established crop mixture, and its evolution within a farmer-led seed circulation network, considering this particular design of on-farm dynamic management as an on-farm evolutionary experiment. Submitted to contrasted environments, seed management and farming practices in terms of field size or sowing date, genetic diversity of crop populations was maintained over time, while multilevel differentiation among populations was detected at the genetic and phenotypic levels highlighting the contribution of genetic drift and selection in the evolution of these crop populations.

Our findings highlight the remarkable ability of the mixture to respond to selection in drastic conditions. While we initially expected population differentiation to be mostly mediated through variation in the proportions of the mixture components, we found that within-component genetic evolution also substantially contributed. In particular, the TSB landrace, the most diversified landrace of the four, was identified as the keystone in the adaptation process of the mixture. This landrace was present in all populations, and it responded with different strategies for earliness depending on global environmental and agricultural conditions. These findings emphasize how critical it is to maintain within-variety genetic diversity. The distribution of crop genetic diversity met in on-farm management is a coproduct from the farmers’ self-organization. Therefore, this study shows that such social organization could potentially contribute to the adaptation of crop biodiversity to climate change. In addition, this short-term evolutionary experiment sets the stage for promising properties of mixtures and confirms the potential of genetic diversity to maintain adaptability and stability in changing environments. This investigation needs to be continued on the medium term to confirm our results.

## Acknowledgements

The authors gratefully thank Élise Demeulenaere for advices during the preparation of the survey. They also thank Audrey

Didier from the national French GenBank at Clermont-Ferrand for providing *ex situ* accessions and the farmers of the Réseau Semences Paysannes for stimulating discussions. This work was supported by a grant from the Institut National de la Recherche Agronomique (France) to M. T. A grant from the Bureau des Ressources Génétiques (now Fondation sur la Recherche sur la Biodiversité) (2005–2007 project) provided money for field and molecular experiments.

## References

- Artoisenet P, Minsart L-A (2014) Statistical genetics in traditionally cultivated crops. *Journal of Theoretical Biology*, **360**, 208–221.
- Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.
- Batagelj V, Mrvar A (2002) Pajek-analysis and visualization of large networks. In: *Graph Drawing* (eds Jünger M, Mutzel P), pp. 8–11. Springer, Berlin, Germany.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2000) *GENETIX 4.02, logiciel sous Windows™ pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR, Université de Montpellier II, Montpellier, France.
- Bellon MR, Hodson D, Hellin J (2011) Assessing the vulnerability of traditional maize seed systems in Mexico to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 13432–13437.
- Bonnin I, Rousset M, Madur D *et al.* (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.
- Bretting PK, Duvick DN (1997) Dynamic conservation of plant genetic resources. *Advances in Agronomy*, **61**, 1–51.
- Dawson JC, Goldringer I (2012) Breeding for genetically diverse populations: variety mixtures and evolutionary populations. In: *Organic Crop Breeding* (eds Bueren ET, Myers JR) pp.77–98. John Wiley & Sons, West Sussex, UK.
- Delêtre M, McKey DB, Hodkinson TR (2011) Marriage exchanges, seed ex-changes, and the dynamics of manioc diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 18249–18254.
- Dreisigacker S, Zhang P, Warburton M, Skovmand B, Hoisington D, Melchinger A (2005) Genetic diversity among and within CIMMYT wheat landrace accessions investigated with SSRs and implications for plant genetic resources management. *Crop Science*, **45**, 653–661.
- 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.
- Enjalbert J, Dawson JC, Paillard S *et al.* (2011) Dynamic management of crop diversity: from an experimental approach to on-farm conservation. *Comptes Rendus Biologies*, **334**, 458–468.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Fu D, Szűcs P, Yan L *et al.* (2005) Large deletions within the first intron in VRN-1 are associated with spring growth habit

- in barley and wheat. *Molecular Genetics and Genomics*, **273**, 54–65.
- Garrick RC, Caccone A, Sunnucks P (2010) Inference of population history by coupling exploratory and Model-Driven phylogeographic analyses. *International Journal of Molecular Sciences*, **11**, 1190–1227.
- Goldringer I, Prouin C, Rousset M, Galic N, Bonnin I (2006) Rapid differentiation of experimental populations of wheat for heading time in response to local climatic conditions. *Annals of Botany*, **98**, 805–817.
- Guyomarc'h H, Sourdille P, Charmet G, Edwards K, Bernard M (2002) Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. *Theoretical and Applied Genetics*, **104**, 1164–1172.
- Gómez O, Blair M, Frankow-Lindberg B, Gullberg U (2005) Comparative study of common bean (*Phaseolus vulgaris*L.) landraces conserved *ex situ* in genebanks and *in situ* by farmers. *Genetic Resources and Crop Evolution*, **52**, 371–380.
- Hartigan JA, Wong MA (1979) A K-Means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**, 100–108.
- van Heerwaarden J, van Eeuwijk FA, Ross-Ibarra J (2010) Genetic diversity in a crop metapopulation. *Heredity*, **104**, 28–39.
- Ingvarsson P (2002) A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution*, **56**, 2368–2373.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- 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.
- Kamada T, Kawai S (1989) An algorithm for drawing general undirected graphs. *Information processing letters*, **31**, 7–15.
- Labeyrie V, Rono B, Leclerc C (2014) How social organization shapes crop diversity: an ecological anthropology approach among tharaka farmers of mount kenya. *Agriculture and Human Values*, **31**, 97–107.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.
- Leino MW, Boström E, Hagenblad J (2013) Twentieth-century changes in the genetic composition of swedish field pea metapopulations. *Heredity*, **110**, 338–346.
- Liu F, Zhang L, Charlesworth D (1998) Genetic diversity in leavenworthia populations with different inbreeding levels. *Proceedings of the Royal Society of London. Series B Biological Sciences*, **265**, 293–301.
- 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.
- Maxted N, Ford-Lloyd BV, Hawkes JG (1997) Complementary conservation strategies. In: *Plant genetic conservation: the in situ approach* (eds Maxted N, Ford-Lloyd BV, Hawkes JG), pp. 15–39. Chapman & Hall, London, UK.
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, **17**, 285–291.
- Mercer KL, Perales HR (2010) Evolutionary response of landraces to climate change in centers of crop diversity. *Evolutionary Applications*, **3**, 480–493.
- Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.
- 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.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Nei M (1987) *Molecular Evolutionary Genetics*, pp. 512. Columbia University Press, New York.
- Pannell JR, Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **53**, 664–676.
- 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**, 476–486.
- Pressoir G, Berthaud J (2003a) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. *Heredity*, **92**, 88–94.
- Pressoir G, Berthaud J (2003b) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity*, **92**, 95–101.
- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rhoné B, Remoué C, Galic N, Goldringer I, Bonnin I (2008) Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. *Molecular Ecology*, **17**, 930–943.
- Röder MS, Korzun V, Wendehake K *et al.* (1998) A microsatellite map of wheat. *Genetics*, **149**, 2007–2023.
- Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. *Theoretical and Applied Genetics*, **108**, 920–930.
- Rousset M, Bonnin I, Remoué C *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.
- 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 of the United States of America*, **105**, 18824–18829.
- Samberg LH, Fishman L, Allendorf FW (2013) Population genetic structure in a social landscape: barley in a traditional Ethiopian agricultural system. *Evolutionary Applications*, **6**, 1133–1145.
- Serpoly E, Dawson JC, Chable V *et al.* (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.
- Sherman JD, Yan L, Talbert L, Dubcovsky J (2004) A PCR marker for growth habit in common wheat based on allelic variation at the gene. *Crop Science*, **44**, 1832.
- Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, **12**, 253–262.

- Somers D, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, **109**, 1105–1114.
- Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, **135**, 367–374.
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
- 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 *et al.* (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.
- Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution*, **42**, 995–1005.
- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wright S (1969) *Evolution and The Genetics of Populations, Vol. 2: The Theory of Gene Frequencies*. Chicago Press, Chicago.
- 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.
- Zhang P, Dreisigacker S, Buerkert A, Alkhanjari S, Melchinger AE, Warburton ML (2006) Genetic diversity and relationships of wheat landraces from Oman investigated with SSR markers. *Genetic Resources and Crop Evolution*, **53**, 1351–1360.
- Zuppinge-Dingley D, Schmid B, Petermann JS, Yadav V, De Deyn GB, Flynn DFB (2014) Selection for niche differentiation in plant communities increases biodiversity effects. *Nature*, **515**, 108–111.

---

M.T. and I.G. designed the study and wrote the manuscript. S.T. was involved in the genotyping of markers in candidate genes, analysed the data and contributed to the discussion. S.J.-P. and N.G. managed the phenotyping experiment. C.R. managed the genotyping experiment. I.G. was the lead P.I. on the grant.

---

## Data accessibility

Raw genotyping and phenotyping data with sampling locations are both accessible at the following: 10.5061/dryad.k5f12

## Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Location of the 14 origins of the 21 samples.

**Fig. S2** Correlation between altitude and pairwise  $F_{ST}$  (VMK) between A03 and each other community ( $R^2 = 0.33$ ,  $P_{value} = 0.015$ )

**Table S1** Diversity indices computed in neutral markers and earliness gene polymorphisms.

**Table S2** Sample size ( $S_{x/y}$ ), demographic size ( $N_{x/y}$ ), temporal variations in allele frequencies ( $\bar{F}_c$ ), effective size ( $N_E$ ), and confidence intervals for the effective size ( $CI_{left}$  and  $CI_{right}$ ) computed for different MDT samples.